Combining *Translation into the Second Language* and *Second Language Learning*: An Integrated Computational Approach

Chi-Chiang Shei

Doctor of Philosophy
University of Edinburgh
2003

Dedicated affectionately to:

Mei-fang Yen
Shuan Shei
Farn Shei

You are great family to have.

Abstract

This thesis explores the area where translation and language learning intersects. However, this intersection is not one in the traditional sense of second language teaching: where translation is used as a means for learning a foreign language. This thesis treats translating into the foreign language as a separate entity, one that is as important as learning the foreign language itself. Thus the discussion in this thesis is especially relevant to an academic institution which contemplates training foreign language learners who can perform translation into the foreign language at a professional level.

The thesis concentrates on developing a pedagogical model which can achieve the goal of fostering linguistic competence and translation competence at the same time. It argues that constructing such a model under a computerised framework is a viable approach, since the task of translation nowadays relies heavily on all kinds of computational tools, whereas the computer assisted language learning framework (including the domain of distance learning) advances at a slow but steady pace, which offers a bridge to connect translation and language learning. The theoretical underpinning of the model is established by relating translation competence to linguistic competence. It is argued that a successful translator working in the area of translating into the second language must also be a competent learner of that language, and the instructions for both are inseparable.

At the practical level, the thesis distinguishes three types of software which are relevant in the current context: the translation workstation (TW) based system, the computer assisted translation learning (CATL) system, and the computer assisted language learning (CALL) system. The first kind of system is based on existing translation aid software such as the well-known category of Translation Memory Systems. Besides being used as a computer environment for translating, the translation memory software can also be used to embed second language teaching concepts. The second type of system is the software that is especially developed for teaching translation AND the foreign language being translated into. In particular, the discussion concentrates on a kind of model referred to here as the Translation Micro World, which is an intelligent tutoring system drawing from pre-edited bilingual corpora built into the system. It is shown that this type of construct is especially useful for building up the translator's idiomatic competence in the target language in which the translator is a learner. The third and last type of software is the computer assisted language learning software which can be adapted to incorporate the element of translation. The idea is to embed translation activities in existing CALL constructs such that translation becomes the primary means for learning the

target language. Thus, by covering the whole range of these three types of translation or language learning software, it is hoped that a curriculum aimed at fostering translators translating into the second language can have a rich repertoire of computer assisted learning tools to draw on.

The pedagogical framework proposed in this thesis has to be practicable in an educational setting. The last part of the thesis thus considers the implementation issues of this framework. It is argued that the traditional syllabus design concepts are not fully transferable to the current pedagogical framework. Traditional syllabi consist of the statements of goals and objectives, learning material based on the objectives, lesson plans including presentations, tasks and exercises contrived around the material, and assessments. The current pedagogical model, however, emphasises the use of computers in the curriculum, and thus the goals and objectives need to be stated in a different way, learning material has to be prepared in different forms, and classes have to be conducted in an entirely different manner and at a different pace. Learning tasks, exercises and assessments all have different meanings in a computerised learning environment.

Acknowledgements

Hearty thanks are given to my first supervisor Dr. Helen Pain for her help and guidance in many ways, from the moment I entered the University, through many difficulties and frustrations, until the very last moments before, during, and after the viva, without which not only this thesis would have been non-existent but I myself would have been much less complete than I am now.

I thank my second supervisor Henry Thompson for meeting with me, giving ideas, correcting part of my thesis, and offering valuable administrative help.

I would also like to thank all members of the AIED group for their various help and support, and all those nice people from this Division or from other UK universities who participated in the "collocation questionnaire". Special thanks must go to Kaska Porayska-Pomsta and Shari Trewin, who kindly responded to my requests for native speaker assistance in revising a test paragraph; and to Tim Willis, who kindly helped me print out two copies of the thesis, have them bound in the library, and deliver them to the Faculty Office.

I can never thank enough my internal examiner, Dr. Jon Oberlander, and the external examiner, Dr. Agnes Kukulska-Hulme, who together conducted such an unforgettable oral examination that I will remember their thoroughness, their gracefulness and their helpfulness for a long time to come. Their suggestions for revision and exhaustive lists of mechanical errors made the thesis so much more readable, though any remaining inadequacy in the quality of the thesis must be my own responsibility.

I warmly thank the University staff, in particular Jean Bunten, Betty Hughes, and other Division and Faculty secretaries, and the DAI and COGSCI computer officers, for their help in administrative or computing matters. Special thanks are to Margaret McMillan, for settling me into the well-equipped Visitor's Office during my visit for viva in November 2002.

Thanks are also due to Professor Delia Davin, Dr. Ruru Li, Dr. Ning Yi and other faculty members from Department of East Asian Studies, University of Leeds, for intending to employ me and for their support in my several months' stay in Leeds.

I also thank Dr. Lan Yu-Su, Chair of the Department I worked with for two years since my return from the UK, for helping me to settle in and providing me with various support. No less am I thankful to Miss Chen Yi-xiu, staff member of the Computing Centre of Chang Jung University, for her valuable support in computing.

Finally, I heartily thank my uncle, Professor Chen Chao-long, M.D. for offering extra sponsorship to me throughout my stay in Edinburgh. No less affectionate thanks are given to my sister Hui-fen for her constant support, as well as to all the

other family members.

This research is mainly supported by a four-year scholarship offered by the Ministry of Education, Republic of China on Taiwan.

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text

(*Chi-Chiang Shei*)

Publications

The following material has been published in connection with this thesis:

C-C Shei and Helen Pain. 2000. "An ESL Writer's Collocational Aid". Computer Assisted Language Learning (CALL). 13(2): 167-182.

C-C Shei and Helen Pain. 2001. "Learning a Foreign Language Through Machine Translation: Focusing on Sentence Stems and Collocations". AI-ED Workshop on CALL: Implementing Intelligent Language Tutoring Systems. May 19-23, 2001. San Antonio, US.

C-C Shei. 2001. "Computational Approach to the Teaching of Translation Methods". Studies of Interpretation and Translation. 6(6): 173-187.

Note of Pronoun Usage

In the thesis, the pronoun *he* or *she*, when not referring to a particular person, is intended to be gender-neutral and refers to learners, teachers, translators, etc. in a general sense.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Translator as second language learner

Language production may be more difficult to manage than language comprehension not only from a research perspective (Harley 1995: 244) but also from a pedagogical perspective. Swain (1985) notes that, in methodologies which emphasise comprehensible input, it is possible for a second language (L2) learner to comprehend input without a syntactic analysis, but such learners often produce only limited utterances, because comprehensible output cannot be generated without syntactic analysis, which the student has little expertise in. More specifically, when learning to read and write in L2, it is relatively easy to comprehend a text in L2 as long as the learner has the correct aids and tools. But when it comes to writing in L2, although the discourse produced by the learner may be comprehensible, few learners achieve the native-like proficiency in terms of writing throughout their L2 learning career: they tend to produce interlanguage at various stages of approximation to the norm of L2, which show inadequacy in grammar or lexis or both.

Another area where learners' ability to produce language is less satisfactory than their comprehension is translation. In the field of translation, translating into the first language (L1) is the preferred mode of working. The preference is easy to understand, since in translation probably the most important thing is to produce idiomatic target language (TL) text, which can be done only if the translator is a native or near-native speaker of the TL. For someone who is a learner of the target language, translation into the second language is often incorrect in some way – be it in grammar, lexis or cultural aspects. The learner can resort to a variety of resources like dictionaries and encyclopaedias in comprehending a difficult source text when translating from L2 into his mother tongue. However, there is less possibility for him to seek help from the same resources when the objective is to produce native-like text in L2 (i.e. when the source language is L1 and the target language is L2), since language production is largely a matter of internal resources.

Admitting that translating into the second language is not the most felicitous working mode, in practice the service is still solicited in many branches of human activities. For example, since English is an international language, documents from many other languages in the world need to be translated into English daily for commercial, educational or other reasons. As there are not so many native speakers

of English who also know, say, an Asian language like Vietnamese, most likely native speakers of Vietnamese who also know English have to undertake the task of translating into English. Among the native speakers of Vietnamese, only a small portion will have native-like ability in English as well, and most of them will be learners of English at different proficiency levels. It is then a fact of life that we have to deal with translators as learners of the target language. It will be helpful, within a translation curriculum, if we could find a way to educate the translator on both translation skills and the language itself while helping him with the translation processes.

Therefore, the thesis is dedicated to finding a feasible pedagogical model conducted within the environment of the computer, which can deal with the teaching of a foreign language as well as translating into the foreign language at the same time. This investigation, by its nature, is also expected to throw some light on the use of translation in language learning, which is a perennial issue in second language learning.

1.2 Second language competence and translation competence

Competence in a language can be variously defined and there are perpetual confusions and debates about this issue in the applied linguistic tradition. The debates involve, mainly, whether to define competence as pure "knowledge of the language" or as "ability to use the language". In the case of defining language as pure knowledge, the objection is that competence should also subsume knowing how to apply that knowledge about language to actual usage. In the case of defining competence as the ability to use language, on the other hand, the objection is that it necessarily involves so many aspects of language that the definition quickly becomes too complex and useless.

Translation competence, which involves two languages, is relatively less explored compared with competence in one language. One could say it is easier or more difficult to define. It is easier to define because the tasks of translation are a very small subset of the whole range of tasks involving language and thus competence in doing so is easier to isolate and study. It is more difficult to define, on the other hand, because it involves two languages and the problem of defining translation competence becomes the union of the problems of defining the linguistic competence of each language involved from certain aspects.

In a translation and language learning combined framework, however, it is easier to define both translation competence and linguistic competence. That is, we focus our attention on the translation tasks and translation problems at hand (i.e. the

translator as second language learner), and explore the intersection of translation competence and linguistic competence in relation to these tasks and problems. Thus we can have a firmer theoretical focus and avoid the inherent problems of defining linguistic or translation competence, and can still gain theoretical insights and practical implications from the limited generalisations.

1.3 Investigating learner corpora

A learner corpus is a database consisting of spontaneous or elicited discourse produced by the second language learner. In a computerised instruction framework, learner corpora are useful for collecting user information in order to make educational software able to respond to user needs properly. For example, if we want to make composition guiding software, we collect learners' writings and analyse them in terms of the points we would like to address and design the software based on our findings, i.e. where the learners' weaknesses and interests are, etc.

In the preparation of this thesis two corpora have been collected: learner English corpus (LEC) and learner translation corpus (LTC). The LEC consists of English compositions written by Chinese students learning English as a second language. The LTC consists of English translations from Chinese done by Chinese students who are also learners of English. From the LEC a data set called the Collocation Error Library has been extracted which is used subsequently in an intelligent tutoring system (ITS) responsible for finding potential collocational errors in students' writing. The LTC, on the other hand, provides the chance to study language transfer, especially the negative transfer where the learner makes a mistake in L2 by applying the L1 rules. While LEC tells us "what" errors are made, LTC gives us some clues "why" some of them are made. Both corpora are useful for this research in terms of providing theoretical support and empirical data.

1.4 Grammar and fixed expressions

In language production two kinds of processing can be recognised – rule based and instance (or memory) based. In rule-based processing the speaker creates novel sentences from scratch, integrating words into well-formed sentences following some innate grammatical rules (Chomsky 1965). In instance-based processing the speaker draws on examples of language usage encountered in the past and reapplies the entries as they are from memory. From this view, speakers often simply patch some "prefabricated" units of words loosely together to form sentences without having to analyse the chunks (Nattinger & DeCarrico, 1992: 32).

When examining the learner corpora, it was found that, for the less proficient L2 writers, grammar and diction are "equally bad". For the more advanced L2 writers, on the other hand, there are very few grammatical errors, and yet the choice of words is still problematic at times. That is, although they have internalised a reasonably adequate set of grammatical rules in L2, the advanced learners do not make word selections as native speakers do. This could be due to their not having memorised enough idioms, collocations and sentence stems as native speakers have.

Expressions like *bark up the wrong tree* and *have sticky fingers* are idioms, which are often metaphoric in nature and have non-replaceable parts. Expressions like *strong tea* and *sorely needed* are collocations, which are more flexible than idioms (for example, *badly needed* is also acceptable) but are more fixed than free combinations of words (for example, *tough tea* and *harmfully needed* are unacceptable). Besides these fixed expressions in English, there is also a huge list of sentence stems or sentence builders like *At the heart of ... is ...*, *Demand for ... is high*, and *With all the..., it should be no surprise that ...* etc. Arguably it is all these fixed expressions, lacking in advanced L2 learners' writings, which make them sound "foreign" or unnatural by native speakers' standards.

1.5 Intelligent tutoring systems and machine translation

Intelligent tutoring systems (ITSs), when used for language learning, are an extension of CALL (computer assisted language learning) software with more advanced features in the areas of expert knowledge, interactivity, user modelling, natural language dialogues, etc. Because of the advances in computer technology and language engineering, the ITS today provides infinite possibilities for optimising language instruction. For instance, an ITS can offer spectrographic analysis of students' voice input and allow them to approximate the correct pronunciation by comparing their own spectrographs with standard ones (e.g. TellMeMore). An ITS can also monitor a student's progress in vocabulary and advance him to the next stage in reading or vocabulary acquisition on appropriate occasions (e.g. Shei, 2001). An ITS is inferior to human teachers in some respects, such as the use of natural language and body language, the possession of encyclopaedic knowledge and the application of such. In other respects the ITS is undoubtedly superior, such as in keeping large amounts of teaching material and student data and in retrieving them in an instant.

Machine translation (MT) is fast gaining importance in the age of information technology. It is already playing a significant role on the World Wide Web helping with multilingual document retrieval and processing. However, MT nowadays is still quite inferior to human translation at and above the level of syntactic analysis.

Between two very different languages like Chinese and English, there is still a long way to go before MT can offer fully automatic and idiomatic translation. However, with what limited capacity MT has, it can still contribute substantially to some areas of multilingual communication, such as web-based information services and the production of drafts for multilingual documents. In the educational setting, MT also has the potential to contribute to language teaching, among other things, although this feature of MT has as yet rarely been explored.

This thesis includes a discussion of how MT technology could be built into an ITS in order to optimally carry out a number of translation learning or language learning processes. In particular, considering the importance of fixed elements in second language learning and the lack of progress in automatic detection of grammatical errors in language technology, the focus of the ITS is on collocations and sentence stems. The use of the ITS is instrumental, however, and the ultimate goal is to make the point that in translating into the second language, since the translator is an L2 learner more often than not, any reasonable translation curriculum needs to consider translation and language acquisition together when designing teaching methodology and organising teaching materials.

1.6 Genre and audience

It is a well-known fact that translation is strongly tied to the genre of text in hand. In the process of translating, the translator needs not only to analyse the surface linguistic structure but also to worry about the genre and register of the text in order to fulfil the requirements not only in meaning but also in style and purpose of the source text in doing the transfer. Likewise, in evaluating a translation, a possible criterion is to see whether the translation has accomplished the transfer of discourse and communication features appropriately for that particular genre, i.e. using the particular register for the desired audience in the target culture.

In this thesis, most of the translation tasks mentioned are within the range of "general translation" as these are mostly examples taken from actual teaching of translation modules. That is, it generally precludes texts which are either very technical (like the translation of quantum physics essays) or very "literary" (e.g. works in literature rich in implications, metaphors, riddles, etc.), since these are taught in other modules not concerned with the author. Within the range of text dealt with here there are, for example, newspaper articles, fables, written jokes, introductory Web pages, etc., although occasionally, a selection of not very difficult literary works are also ventured into. This thesis does not profess coverage of a wide range of text and the teaching programs or methodologies proposed are thus limited in some way.

1.7 Structure of the thesis

While this chapter introduces and motivates the thesis, the rest of the thesis expands on the ideas explained in this chapter.

Chapter Two explores the issue of 'translator as second language learner' in more depth.

Chapter Three considers the distinctiveness of and relationship between linguistic competence and translation competence and their pedagogical implications.

Chapter Four raises the findings from analysing the learner corpus and the translation corpus, in particular the error types discovered from an error analysis of the corpora.

Chapter Five goes on to explore the significance of the learners' errors in L2 writing and L1-to-L2 translation, from the perspectives of both rule-based and instance-based modes of processing, though the focus is on the latter.

Chapter Six looks at some recent advances in language technology which have a strong bearing on translation or language learning and therefore have strong implications for the pedagogies of such.

Chapter Seven through Chapter Nine each deal with a type of translation or language learning related software which might be considered in the combined approach of translation teaching and foreign language teaching.

Chapter Seven introduces the translation memory system, which is a kind of ready-made translation aid software. The discussion centres around the role of such software in a translation curriculum, and how it can incorporate the aspects of language teaching.

Chapter Eight introduces a kind of software which might be called the "Translation Micro World". This is a kind of intelligent tutoring system in which language corpora play a significant role. The idea is to offer a controlled environment (in terms of the range of language used, etc.) where the student can shape his language according to the norms specified by the instructor or illustrated by the inbuilt corpus.

Chapter Nine in turn looks at the possibilities of integrating the elements of translation into CALL (computer assisted language learning) programs in the ordinary sense (i.e. usually not referring to or making use of translation).

Following this series of discussions on the categories of software which are potentially usable in a language/translation learning combined syllabus, we go on to discuss in Chapter Ten how a syllabus can be contrived for such a course which relies heavily on the use of these types of software.

Chapter Eleven, on the other hand, addresses the issues of classroom implementation and user feedback.

Finally, Chapter Twelve gives a summarisation and conclusion of the whole investigation, relating the findings to the fields of both translation teaching and language education.

Chapter 2

Translation into the Second Language

Both translation practice and translation theories have a long history. The focus of this thesis is inevitably concentrated on those of the contemporary era, as this research is strongly associated with modern theories and practice of linguistics, applied linguistics and computational linguistics and does not touch upon the literary or philosophical side of translation. Consequently we will refrain from discussing the translating of the Bible, for example, and the views of John Dryden (at least not directly) or Quine on translation etc. in this thesis.

The nature of this thesis is largely pedagogical and practical in the sense that it tries to find an acceptable and efficient model for combining the instruction of language and translation under a computational framework. The theories touched upon are those which have a direct bearing on establishing or validating the means for achieving this end. This chapter first introduces the kinds of translation work done today to give a general picture of the practical side of translation at the time of writing. Then it reflects on the nature of translation theories and points out the theoretical backgrounds relevant to this thesis.

As the focus is on the individuals doing the translation (i.e. native speakers or language learners), the chapter spends another two sections explaining the differences between translating into the first language and the second language, and considers why the former is preferred to the latter, and why nevertheless translation into the second language is still justifiable and needed.

2.1 Translation practice

According to Sofer (1998), there are three main sources of translation work in the United States: the government, the public sector and the private sector. The government sector requires translation from the local government's pamphlets and brochures to the federal government's documents related to world trade and international relations. The public sector includes all kinds of organisations, foundations and academic institutions which need translation services. The private sector refers to the translation agencies which act as brokers between work sources and the translator.

In the case of Taiwan which is of particular interest to this thesis, in the city of Taipei alone there are more than a hundred translation agencies. While the sources

of translation work are the same as those of the US, the nature of the translation work can be found through a survey of available Web pages of translation agencies. The primary functions of the translation companies in Taiwan are as follows:

- Document translation: Documents which require translation range from technical writing, business letters, advertising, medical reports and legal documents to journal articles and dissertations.
- Interpretation: Interpretation is a different discipline from translation. It renders the target language in the oral rather than written form. This usually includes conference interpretation (consecutive or simultaneous), business interpreting, diplomatic missions, court proceedings, etc.
- Desktop publishing: It is unlikely that a client would ask for a translation agency to translate a document and take the translated document, say, to a printing company to publish a brochure. Modern translation companies are equipped with desktop publishing software like PageMaker, Photoshop etc. to combine texts and pictures into a document for high-quality printing.
- Video production: This includes dubbing and subtitling of videos and films, computer games, etc.
- Software localisation: Localisation is necessary for the globalisation of business. Software localisation involves not only translation of texts but also compiling, debugging, resizing etc. of the software itself.
- Website localisation: This usually happens in the form of entire website project planning, translating from existing sites or from scratch, covering HTML scripting, CGI programming etc.
- Application packages: In Taiwan there are a large number of people applying for overseas immigration or school study daily who have to rely on the translation agencies to translate the necessary documents and organise the application forms etc. Often an entire application package is offered at a lump sum.

As can be seen from the above categorisation of translation businesses, the translation jobs being done nowadays are very different from those at the time of Bible translation.[1] This echoes what Hatim et al. (1990) say: "Few can earn a living from literary translating, which is often done for the sake of the intellectual or aesthetic pleasure it procures." (p. 20). Likewise Sofer (1998) comments:

[1] Or other more recent times when literary translation was the focus. During the cold war the US spent a great deal of time and money on Russian to English translation (Pain 2002).

> High quality literary translation has always been the domain of the few, and is hardly lucrative ..., while technical translation is done by a much greater number of practitioners, and is an ever-growing and expanding field with excellent earning opportunities. (p. 37)

The implication of this proliferation of translation work in business and industry is far-reaching. For one thing, this encourages the gradually heavier reliance on machine translation and the use of computers as translation aids (Hutchins, 1999). But what is more relevant to us is the large demand for the practice of translating into the second language in local translation markets where the number of native speakers of the second language falls far short of demand. For example, a Taiwanese family applying to immigrate to Canada needs the translation company to translate or write relevant English letters for them. There are not that many native speakers of English living in Taiwan doing translation as a job. (For the handful of native speakers living here teaching English is a much more popular and lucrative option) More likely the employees in Taiwanese translation companies are learners of English as a foreign language, i.e. Taiwanese natives, rather than native speakers of English. The documents translated from Chinese to English are thus written by learners of English -- but this practice is apparently surviving well because the translation business in Taiwan is quite alive and thriving.[2]

2.2 Translation studies

Although the practice and study of translation have had a long history, it was only relatively recently that translation started to be treated seriously as an independent academic discipline. Two of the most well-known precursors in modern translation research are Nida, for his scientific treatment of Bible translation (Nida 1964), and Catford, who proposed the term "translation shifts" to explain the process of translating (Catford 1965).

The name *Translation Studies*, according to Bassnett (1991), was first proposed by André Lefevere in 1978 "for the discipline that concerns itself with 'the problems raised by the production and description of translations'" (p. 1). Baker (1998), on the other hand, accredited the term to James Holmes, who "argued for the adoption of 'translation studies' 'as the standard term for the discipline as a whole'" (p. 277). One way or another, the appearance of the term seemed to have a consolidating power and now *Translation Studies* is widely adopted as a cover term for translation related research.

---

[2] The first academic institution of translation was established in Taiwan in 1988. In 2000 there are four such institutions in Taiwan, reflecting the increasing demands for translators in the industry.

Probably the most significant thing about translation studies is its interdisciplinary nature. Bassnett (1991), for example, makes the following observation:

> A distinguishing feature of work in Translation Studies has been the combining of work in linguistics, literary studies, cultural history, philosophy, and anthropology. (p. xi)

Neubert & Shreve (1992) also claim:

> The development of translation studies over the last two decades has been characterized by a decline in the influence of linguistics and a movement to give translation research an interdisciplinary focus. (p. viii)

A positive consequence of this interdisciplinary nature is that each scholar can approach the issues in translation from a particular perspective and gain some useful results -- what Neubert & Shreve refer to as 'selective attention' (ibid.).

Many researchers in translation also highlight the empirical nature of translation theories. Newmark (1988) says concisely, "The heart of translation theory is translation problems" (p. 21). Bassnett (1991) further explains, "The need for systematic study of translation arises directly from the problems encountered during the actual translation process" (p. 7). Even more forcefully, Bassnett goes on to say, "To divorce the theory from the practice, to set the scholar against the practitioner as has happened in other disciplines, would be tragic indeed" (ibid.). What do we do with the practice of translation? Neubert & Shreve (1992) explain, "Our task is to recognize in the infinite variety of the practice of translation significant patterns and regularities" (p. 34). This brings us back to what Newmark says, i.e. that "translation theory broadly consists of, and can be defined as, a large number of generalizations of translation problems" (p. 21).

In this respect the study of translation is much like linguistics itself. From the 1960's onward the mainstream Chomskyan tradition pursued linguistics on a highly theoretical basis (so-called 'armchair linguistics'). Though discourse analysis, functional grammar and psycholinguistics continue to have their advocates and followers, the usefulness of empirical linguistic data have never been so compelling and convincing than that illustrated by corpus linguistics from the 1980s onward. It can only be natural for translation studies, so closely related to linguistics, to follow the same path. The use of bilingual or parallel corpora to compare language pairs and to uncover the nature of translation is already prevailing. As Kenny (1998) says, translation studies "makes very particular demands on corpora", and "corpora are already leading to new ways of looking at translation".

Echoing this empirical nature of translation studies, historically there have been some issues in the practice of translation which are the central focus of scholars' attention. Hatim et al. (1990) conveniently summarise issues and debates in translation that have haunted scholars now and in the past, some of which are:

- Literal versus free: This is a "time-honoured debate", in Hatim et al's words, which is concerned with the degree of latitude the translator is allowed in translating the source text into the target text.
- Process and product: What is available for study is usually the target text itself -- the end product of translation. But what is less accessible -- the processes and procedures of translation -- may be more important to the researchers.
- Objectivity or subjectivity: This concerns the judgment of the quality of translation. Objectivity is ideal but subjectivity is inevitable. Hence, "serious attempts have been made to establish translation criticism on a proper footing" (p. 4).
- Author-centred and reader-centred: Should the utmost concern of the translator be the loyalty to the source text author or to the target text reader? The actual translation produced will vary according to the attitude taken.

Apart from the 'issue oriented' theories of translation, there is also a 'translation model' approach to translation studies. Newmark (1988), for example, proposes a model of translation consisting of four levels:

- The textual level: Roughly, the literal translation level.
- The referential level: The factual level concerning 'encyclopaedia knowledge'.
- The cohesive level: The level "linking the first and the second level".
- The level of naturalness: A set of requirements for the target language used.

Neubert & Shreve (1992), who highlight the interdisciplinary nature of translation studies, go on to suggest several models of translation: the critical model, the practical model, the text-linguistic model, the sociocultural model, the computational model, and the psycholinguistic model. They claim that "without yielding their specific perspectives, each of these models could contribute to a more ambitious and more adequate integrated theory of translation" (p. 32).

For the present study, translation theories are relevant only to the extent that they shed light on the processes of translation in a way that is informative to the study of language acquisition. In other words, we are interested in the intersection of translation studies and second language acquisition theories. This is not attempting

to go back to the Grammar-translation period where translation was used as a means for second language instruction. In that era the theory was prescriptive and deductive, and at present, it is descriptive and inductive. In fact, during the course of this thesis, we will be gradually building up a model combining translation studies, second language acquisition and computer-assisted language learning methodologies.

2.3 Translating into mother tongue

Researchers and practitioners of translation often regard translating into the mother tongue as the more, if not the only, acceptable working mode of translation (as compared to translating into the second language). Sofer (1998), for example, explains:

> Generally speaking, one translates from another language into one's own native language. This is because even years of study and experience do not necessarily enable one to be completely at home with an acquired language. (p. 34)

Newmark (1988) is also suspicious of someone who translates into a target language without being a native speaker of it. He claims that when one is doing a 'communicative translation' (as opposed to 'literal translation', for example),

> ... whether you are translating an informative text, a notice or an advert, 'naturalness' is essential. That is why you cannot translate properly if the TL is not your language of habitual usage. (p. 26)

In practice, the attempt to maintain this 'native speaker standard' in translation can also be seen in the advertisement of translation companies' Web pages. For example, one such Web site says, "All of BtChina's translations are handled by professional native speakers";[3] while another emphasises that their translated documents "read as if they had been originally written in the target language".[4] Another translation company in Taiwan says to their potential customers, "It is better, therefore, to trust your translation work to native-speaking translators than to foreign language professors" because "even a Sinologist with a Ph.D. may not speak Chinese as fluently as a native high school student in China".[5]

What constitutes a native speaker of English seems clear, especially in the United Kingdom, the United States, Canada, Australia, and New Zealand. However, for

---

[3] BtChina, at: http://www.btchina.com/.
[4] Boffin China, at: http://www.boffin.com.cn/.
[5] American Translation Link, at: http://www.translation.com.tw/.

people who live in a country other than those mentioned above which also uses English as an official language (e.g. India and Singapore) who grow up speaking English, are they native speakers of English or not? They are because English is one of their first languages. They are not because their English will differ from the 'standard English' in some way. This is the problem of 'World English' which we will not go into here due to the scope of the thesis.[6] Suffice it to say that who are the native speakers of English and what is standard English are sometimes hard to define. This leaves some room for the usability and justifiability of the practice of translating into the second language. That is, there may be some fuzzy areas in the definition of 'English native speakers' who translate from other languages into English, or in the English they write.

2.4 Translator as second language learner

Anyone who does not speak or write a second language as native speakers of that language do can be put into the broad category of second language learners. This includes both students all over the world learning English as a foreign language (EFL) as beginning or intermediate learners, and advanced learners like their teachers who are not native speakers of English, such as the author of this thesis himself.

A learner's knowledge of the second language is sporadic rather then systematic. Being a learner of a language means that the language produced by such a person is always a kind of 'interlanguage' (Selinker, 1969), which is on an evolutionary path to the norm of that language. This means the target language produced by the learner is always "defective", or incomplete, in some way; that is, it does not correspond to the language normally used by adult native speakers and deviates in the area of syntax, lexis, phraseology, text structure, semantics or pragmatics. However, not all errors in these areas will constitute barriers for communication between native speakers and non-native speakers.

According to Ellis (1994), non-native speakers, when asked to judge the quality of learners' English, are more inclined to notice the irregularities with respect to 'form'. Native speakers, on the other hand, tend to pay more attention to the errors which interfere with the communication of meaning (pp. 66-67). Thus theoretically there should be no intrinsic hindrance for EFL learners to produce English translations acceptable to native speakers as long as they are competent in terms of communication.

There is a significant disadvantage, however, which makes translation into the

[6] Nunan (1999), for example, raises questions about the standard of English and the distinction between first language and second language.

second language even more difficult than plainly communicating (writing or speaking) in the second language. In translating, the learner is further restricted by the source language text and cannot adopt the 'avoidance' strategy which is available in speaking and writing, where the learner can choose the expressions he has confidence in and avoid uncertain expressions. By being given a set of source texts in translation, however, the 'raw materials' are fixed, so to speak, and the translator's task is to 'match' rather than to 'choose'. And if there are no internalised materials (i.e. expressions in the target language) in the learner's head to represent the source text, then the translation may be doomed to fail miserably.

Nevertheless, as Campbell (1998) comments, "for many parts of the world, translation into the second language is a regular and accepted practice" (p. 12). Earlier we talked about models of translation. Campbell's model of 'translation competence' is one of the rare translation studies which deals exclusively with translation into the second language. We will go into Campbell's model in more detail in the next chapter. For now we simply note that, just as Campbell's task is "to produce a model of translation competence that will serve a pedagogical purpose" (p. 18), the goal of this thesis is to serve the same pedagogical purpose, with a slightly different model -- one within a computer-assisted language learning framework aiming to improve the competence and performance of the translator as second language learner.

2.5 Summary

This chapter first introduces the practice of translation in the professional setting nowadays. It is noted that technical translation is gaining more importance while the status of literary translation is degrading at the practical level. Echoing this empirical emphasis is the discipline called 'translation studies', which is also briefly introduced. The distinction is then made between translating into mother tongue and translating into the second language. It is argued that the latter, though not a standard practice as the former, is nevertheless an acceptable and valuable service. The entire thesis, in fact, is devoted to finding a computational model for supporting the practice of translating into the second language.

Chapter 3

Translation and Language Learning

In order to find a good way to combine translation and language teaching together in the same curriculum, we have to start from the core of their common ground. First we need to know what it means to know a language and to know how to translate into a language. It is thus essential to investigate what constitutes one's knowledge of a language, including the declarative facts about the language, as well as the procedural ability to apply that language to situations. Then we should also ponder on what translation competence consists of and what its relations to linguistic competence are. Equipped with this knowledge, we should then familiarise ourselves with the language teaching and the translation teaching methodologies separately, to know how researchers and practitioners apply the results of research to the pedagogical setting. Finally, we shall consider the existing methods for incorporating translation into language teaching or vice versa, before proposing a model of instruction of our own.

3.1 Competence and performance

We will discuss the relatively well-explored notion of competence in a language first, before looking into the less researched area of translation competence which involves two languages. Inevitably, we wish to find some connections or overlapping areas between these two if possible, so that our combined model of language teaching and translation teaching (to be proposed later) can be informed accordingly.

3.1.1 Linguistic competence

The desire to separate competence from performance in studying languages was first expressed by Chomsky in the following way:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its (the speech community's) language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance (Chomsky, 1965:3).

This aroused notorious debate which centred around whether it is possible to study

language in such a 'purified' environment envisaged by Chomsky; that is, is language still language without being used? Cook (1996) describes the dispute as a battle "waged as much in language teaching, in children's languages, or in computational linguistics, as it is in linguistics itself". For some theoretical linguists, the separation of language forms from their context of use may be desirable, but for applied linguists, obviously language should not be learned as an abstract and isolated system.
Allen (1975), for example, reflects on the issue from the perspective of textbook compilation:

> The majority of language teaching texts contain comparatively few examples of completely abstract sentence formulas on the one hand or of 'authentic' utterance-tokens on the other. Most classroom material consists of sentences which fall somewhere between the two extremes. (p. 40).

That the applied linguists are drawn into the competence-performance contention is inevitable, since it is important to recognise the nature of native speakers' competence in a language so as to set the goals of second language instruction. Stern's (1992) caution that "our concept of competence has a bearing on syllabus design and evaluation at the policy level" (p. 74) amply describes the importance of the task.

It is unimaginable that most applied linguists would be satisfied entirely with Chomsky's definition of *competence* since this cannot serve as the goal of language instruction. We obviously cannot teach students the abstract forms and rules of language *only* and expect them to deal with the performance issues all alone. As Spolsky (1972) says, "this [linguistic competence] is not enough for practical or educational purposes; we are interested not just in the fact that someone knows a language but that he knows how to use it". The mapping of language forms to use does not come from nature but is learned from experience, and language instruction should facilitate this process.

Due to the difficulty of applying Chomsky's concept of competence directly to language teaching, Hymes' later formulation of 'communicative competence' seems more acceptable to the applied linguistics field. Stern (1992) explains:

> Hymes (1972) argued that in addition to linguistic competence the native speaker has another rule system. That is, he knows intuitively what is socially appropriate or inappropriate and can adjust his language use to such factors as the topic, situation, and human relations involved. (p. 73).

Widdowson (1989) also observes that "Hymes proposed his concept of communicative competence in reaction to Chomsky, and it is customary to present it as an improvement in that it covers aspects of language other than the narrowly

grammatical.".

However, Hymes proposes the notion of communicative competence from a sociolinguist's perspective, and it is Canale & Swain (1980) and Canale (1983) who later expand it to form a model of second language teaching in which the communicative competence consists of four parts: grammatical competence, discourse competence, sociolinguistic competence, and strategic competence. Grammatical competence includes phonology, vocabulary, syntax and semantics. Sociolinguistic competence consists of sociocultural rules having to do with language use. Discourse competence refers to rules of discourse like cohesion and coherence. Strategic competence is the ability "to compensate for breakdowns in communication" and "to enhance the rhetorical effect of utterances" (Canale 1983:339).

The addition of communicative competence, and the elaboration of it, does not solve the problem of how to pin down the goals for second language instruction, however. Widdowson (1989:134), for example, raises questions about the nature of the 'ability to use language' since "[a]s soon as you talk about competence as *ability*, or what people can actually *do* with their language, you get into all kinds of difficulty", the reason being that "there is *so much* you have to allow for in the way of individual differences, varying circumstances, attitude, and so on that specification becomes impossible" (italics original). Widdowson's position is that "grammar needs to be in its place" and that we should "allow for the rightful claims of lexis" since, as he believes, the actual use of language may depend more on stocks of lexical items rather than the analysis of structures.[7] The implication is that so-called grammatical competence may have very little to do with the actual use of language at all and it may be wise to still keep competence in the form and that in the use of language apart.

Lyons (1996), on the basis of the assumption that there is a psychological difference between prepositional (or declarative) knowledge ("knowing that something is or is not so") and performative (or procedural) knowledge ("knowing how to do something"), discusses what kind of knowledge linguistic competence is. Lyons emphasises the incorrectness of the conception that "knowledge of a language of the kind that issues in performance is knowledge about the language [sic]". Judging from his "uncontroversial non-technical definition" that "linguistic competence is the knowledge of particular languages, by virtue of which knowledge those who have it are able to produce and understand text in those languages" (p. 16), we can say that Lyons seems to consider the ability to use language as a kind of

[7] A point very important to this thesis (though not crucial in this chapter) which will be revisited in the next chapters.

procedural knowledge, like being able to ride a bicycle without being able to write down the rules of muscle movements etc. effected in riding, and this ability seems to be integrated with knowledge of the language or to come naturally as a result of knowing that language. Lyons apparently disagrees with the idea that man consciously 'knows' the language he internalises and then consciously 'applies' some kind of rules having to do with manipulating the meta-knowledge when using the language in specific situations.

In a way Brown (1996:202) offers a positive and pedagogically friendly conclusion to the competence-performance debates from a "virtuous spiral" perspective. She envisages:

> Initially there is a competence which consists of fundamental constraints on the nature of human languages. Then the child's perception of the language which he or she hears (performance) begins to modify competence, and thus the process continues, with more and more performance (comprehension and production) leading to further modifications in competence, which in turn leads to further modifications in future performance.

What concerns us directly here is Brown's later remark that "just as performance modifies and restructures competence in the first language, so it modifies and restructures competence in the second language". In other words, Brown seems to think that the basic language acquisition mechanisms for the first language and the second language work on the same principles. In Brown's model, competence is an undifferentiated language faculty permanently residing in human brains, though it keeps being added to and subtracted from by performance factors brought about by the production and comprehension of language. This could serve as useful concluding remarks to the competence-performance debate. However, to inform second language teaching we need a more specific model to work on. As Stern (1992) observes, "there has been a strong movement in language pedagogy towards such closer specification of language learning objectives" (p. 77).

Bachman (1990) offers a hierarchical model of language competence. In this model, language competence is first distinguished into two broad categories of organisational competence and pragmatic competence. Organisational competence consists of grammatical competence and textual competence; whereas pragmatic competence contains illocutionary competence and sociolinguistic competence. Let us examine these components of language competence in a little more detail:

- Grammatical competence: This refers to "the knowledge of vocabulary, morphology, syntax, and phonology/graphology" (p. 87).
- Textual competence: This "includes the knowledge of the conventions for joining

utterances together to form a text" (p. 88).

- Illocutionary competence: According to Bachman, this could be explained either by the speech acts theory (Austin 1962) or in terms of language functions (Halliday 1973, 1976). These are all concerned with how we use words to do things in the world.
- Sociolinguistic competence: Any language function is carried out in a specific context with its unique sociocultural and discursive features. Sociolinguistic competence "enables us to perform language functions in ways that are appropriate to that context" (p. 94).

Bachman's model seems to cover the aspects of second language teaching quite well. If there is something left out, this would be the ability to use collocations and other prefabricated units of word combinations which seems subsumed neither by the grammatical competence nor by the textual competence. One likely solution would be to change grammatical competence to lexical-grammatical competence, which would then correctly exhibit at least equal importance of lexis with grammar in one's language competence.

3.1.2 Translation competence

In contrast to the competence in a single language, competence in translation is a relatively much less explored area, at least not in the applied linguistics field using the research terms and convention specific to the discipline.

A recent study on translation competence using the applied linguistics research methodologies was conducted by Campbell (1998). Campbell investigated the translation competence of non-native speakers translating from their mother tongues into English. Among the subjects studied were native speakers of Arabic studying in the translation and interpretation department of an Australian university. Based on the analysis of translation data collected from examinations and translation exercises, Campbell proposed a model of translation competence consisting of three parts:

- Textual competence: Translators have good textual competence when the target texts they write "have the structural features of formal, written English" (p. 73). Translators have poor textual competence "when their output resembles informal spoken English" (ibid.). The structures used for evaluation are nominalizations, type/token ratios, word length, passives, prepositional phrases, etc.
- Disposition: In terms of the strategies for choosing words in constructing target texts, Campbell develops two parameters for describing translators' behaviours –

1) persistent versus capitulating (Does the subject endeavour to translate every word in the source text or does he omit problematic items) and
2) prudent versus risk-taking (Does the subject use common translation equivalents or does he use unusual ones).

The two parameters yield four combinations of disposition in translating:
1. Persistent and risk-taking, 2. Capitulating and risk-taking, 3. Persistent and prudent, 4. Capitulating and prudent.

- Monitoring competence: There are two facets to this part -- self-awareness and editing. Campbell finds better translators to have higher awareness of the quality of their own output as well as more effective editing strategies. Inferior translators tend to overestimate their own performance and use ineffective editing strategies.

Campbell's model gives the feeling that it is customised according to the dataset he has collected and the analysis he is able to conduct based on these data. It thus seems not comprehensive enough to account for the full range of translation competence. For example, the non-trivial aspect of source text to target text correspondence is not investigated, whether from the structural, the semantic, the pragmatic, or the cultural level.

A more comprehensive model of translation competence can perhaps be inferred from practical (as opposed to research-oriented) work like that of Sofer (1998). Sofer proposes ten requirements for professional translators: (pp. 33-37):

1. A thorough knowledge of both the source and the target languages.
2. A thorough "at-homeness" in both cultures.
3. Keeping up with the growth and change of the language and being up-to-date in all of its nuances and neologisms.
4. Always translating from another language into one's native language.
5. Being able to translate in more than one area of knowledge.
6. Having the facility for writing or speaking and the ability to articulate quickly and accurately, either orally or in writing.
7. Developing a good speed of translation.
8. Developing research skills, being able to acquire reference sources essential for producing high quality translation.
9. Being familiar with the latest high-tech developments.
10. Always checking to see what kind of a potential one's language specialty has in a given geographic area.

The above ten points are actually Sofer's "essential criteria for developing a translation career" (p. 36). Though most of them refer to the translator's personal ability ranging from linguistic competence to cultural awareness to technological skills, some of them are pragmatic concerns rather than the inherent properties of the translator (e.g. points 4, 7 and 10). If we compare these with Campbell's model, we find that the two models seem to compensate each other: Campbell offers a good account of how translation competence can be decomposed into manageable linguistic features; whereas Sofer explains in a more comprehensive and heuristic way what temperaments and abilities a modern translator should have.

3.1.3 Competence in translating into L2

It is time to ponder the competence issue at the intersection of second language learning and translation. The objective is to know what constitutes the part of competence for a second language learner which is also applicable to doing translation into that language. We can then concentrate our effort on developing a pedagogical model in this area where linguistic and translation competence overlap.

Recall from our discussion in 3.1.1 that a reasonable conclusion about the competence-performance debate is that competence constantly develops with the modification of performance (Brown, 1996). Since competence is intermingled with performance, from the second language didactic perspective it should be safe for us to not probe too deeply into abstract linguistic systems, and to investigate only the pedagogically tangible, while still being able to say that we are dealing with linguistic competence at a certain level.

Thus we can conflate Canale's (1983) model and Bachman's (1990) model to get the following features of second language competence:

- Lexical-grammatical competence (This author's modification of both Canale's and Bachman's grammatical competence)
- Textual competence (Bachman's subsuming Canale's discourse competence)
- Illocutionary competence (Bachman's)
- Sociolinguistic competence (Bachman's and Canale's)
- Strategic competence (Canale's)

Likewise, we conflate Campbell's (1998) model and Sofer's (1996) criteria to get the following model of translation competence:

- Textual competence (Campbell)

- Lexical competence (Campbell's "disposition")[8]
- Monitoring competence (Campbell)
- Cultural competence (Sofer 2)
- Self-updating competence (Sofer 3)
- Content-knowledge competence (Sofer 5)
- Velocity competence (Sofer 6, 7)
- Research competence (Sofer 8)
- Hi-tech competence (Sofer 9)

A first look at our two conflated models, the first for second language competence and the second for translation competence, immediately calls our attention to the two overlapping modules -- the textual competence and the lexical-(grammatical) competence. The sociolinguistic competence from the language model and the cultural competence from the translation model are also essentially the same and can be subsumed under the name of 'sociocultural competence'. Thus we can perhaps safely say that the intersection of these two models, that is, the competence for a translator as second language learner, consists of the following three core modules:

- Lexical-grammatical competence
- Textual competence
- Sociocultural competence

A less obvious connection between the linguistic and the translation models is the strategic competence of the former and the velocity competence of the latter. According to Canale & Swain (1980), strategic competence refers to

> ... verbal and nonverbal communication strategies that may be called into action to compensate for breakdowns in communication due to performance variables or to insufficient competence" (p. 30).

In this sense, strategic competence seems more appropriate for describing ability in dealing with spoken interactions and so is not directly relevant to translation, which dwells mainly on written language. However, we can readily extend the concept of "compensating" to the process of translation to cover the translator's strategies in anticipating reader's incomprehension and effecting remedial measures beforehand.

---

[8] Note Campbell's translation competence regarding disposition operates mainly on the lexical level, i.e. whether a translator is persistent in attempting to treat every word in the source text, and whether he is willing to take risks in lexical choices or not.

It is then feasible to use "strategic competence" as an cover term for tackling efficiency issues in both mono-language use and translation activities, with more emphasis on the translation side, given the main focus of this thesis.

As for the other components which belong to the union but not to the intersection of the two models, first the illocutionary competence in the language model cannot be found in the translation model anywhere. In fact, the theory of speech acts can be related either to psycholinguistics (e.g. Levelt, 1989), i.e. from the perspective of human language processing – from intention to articulation, or to sociolinguistics (e.g. Wolfson 1989), i.e. from the perspective of social-cultural context or with an interpersonal focus. Here it seems appropriate to allow Bachman's illocutionary competence to be subsumed in his sociolinguistic competence when we are dealing with translation, which is essentially an intercultural and interpersonal business. Next, Campbell's monitoring competence seems more like an editing skill rather than an inherent competence. Sofer's four other criteria -- self-updating, content-knowledge, research and hi-tech competence -- are even farther away from the linguistic model and belong exclusively to the translation model. These should then occupy a peripheral part in our model. Altogether, we have a model of competence specifically for translator as second language learner, as in Figure 3-1.

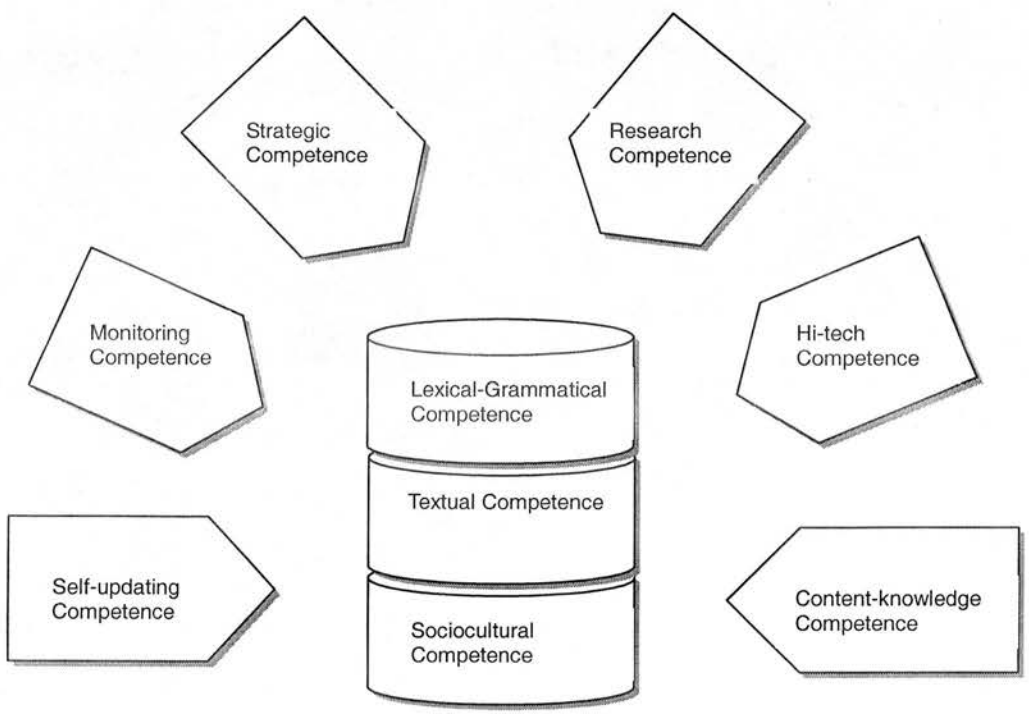


Figure 3-1: A Model of Translation Competence for Second Language Learners[9]

[9] Note the peripheral components of this model largely follow Sofer (1996), which is a practical guide for would-be translators. This reflects the application-oriented nature of the model. In this sense,

The model in Figure 3-1 shows three core components of translation competence surrounded by six peripheral components. Comparable to Brown's (1996) account of the "virtuous spiral" in second language acquisition formed by the evolving competence and the modifying performance, the second language learner as translator in our model also starts with a core competence made up of his knowledge in the lexical-grammatical, textual and sociocultural domains in the target language. But in this case, the learner's competence in translating into the second language does not consist of a self-contained linguistic module (the three core components) only. Rather, the language competence needs to be constantly informed and updated by the translation module (the six peripheral components), apart from being modified by the purely linguistic performance factors described in Brown's cycle.

3.1.4 Summary and implications

So far we have done some analytical and synthesising work, the aim of which is to figure out, from a theoretical point of view, what might constitute an L2 learner's competence in translating into the second language. Our main line of explication is that there are two sets of competence features, so to speak, one for the second language as in the competence of a second language learner, and one for translation as the translation competence of a fully bilingual translator. For a trainee translator who is learning the L2 and learning to translate into the L2 at the same time, however, the two kinds of learning activities can no longer be considered separately. A pedagogical model must be developed which exploits the best possible methodologies for dealing with this kind of learning.

From the above discussion it should be clear that, although competence in the L2 and competence in translating into the L2 can still be distinguished on the part of the learner involved, the boundary is not clearly cut. In fact, the two subsets of competence are closely interrelated and they help each other grow during the translation and L2 learning activities. Thus in the area of teaching translation into a second language, a helpful pedagogical model should be one that bears consequences on acquisition of the second language even though the overt objective is related to translation. On the other hand, L2 learning pedagogies can also consider translating into the L2 as a viable means for enhancing L2 learners competence in the target language.

---

the peripheral components are changeable following the most current trends in the practice of translation.

3.2 Language teaching

In this section we are mainly concerned with the theories and practice of modern second language teaching. Some language teaching methodologies are discussed in the first section with special reference to the status of translation in each approach. In a subsequent section, some research findings in SLA (second language acquisition) are discussed which support the use of translation in second language teaching.

3.2.1 Teaching methodologies

The aim of this section is to find out how different second language teaching methodologies utilise translation, covertly or overtly, as a means for helping learners to achieve second language proficiency. A point of caution here: since the 1990s, the second language teaching profession does not seem to have favoured a particular methodology (as in the fad of communicative language teaching in the seventies/eighties). Instead, theories and techniques in language instruction tend to be discussed independently and applied to teaching without being affiliated to a particular school of didactics (See Kumaravadivelu, 1994). Therefore discussing the place of translation in second language teaching based on methodologies is convenient but it does not include all possibilities of utilising translation in language instruction. Furthermore, only a selection of five prominent methods or approaches are mentioned here to give an illustrative rather than exhaustive picture of how translation has been used in language teaching.

1. GRAMMAR-TRANSLATION

The Grammar-translation method is one of the most influential and persistent approaches in the history of second language teaching. According to Richards & Rodgers (1986: 4), "Grammar Translation dominated European and foreign language teaching from the 1840s to the 1940s, and in modified form it continues to be widely used in some parts of the world today." Grammar-translation is also arguably the methodology which makes the most use of translation in second language instruction. The method gets its name because it approaches the target language first by analysing its grammatical rules and then by applying the knowledge of grammar to the task of translation. As for the direction of translation, Celce-Murcia (1991) comments that the typical exercise in this approach is "to translate sentences from the target language into the mother tongue". But Richards & Rodgers (1986: 4) also claim, "Much of the lesson is devoted to translating sentences into and out of the target language"; that

is, translation into the second language is also possible or, in some cases, required. In either case, translation is a means for enforcing "the vocabulary and grammar encountered in the current and earlier units" (Cook, 1998). The translation activities in the Grammar-translation method are usually sentence-bound and context insensitive because of the method's inherent focus on structure. This differs from real translation tasks or even from genuine translation practices as in a translation class, which are usually meaning-based and focusing on whole paragraphs.

The Grammar-translation method came under attack at one time, mainly on the grounds that it neglects the spoken language, it uses artificial sentences to illustrate grammatical rules, and it relies too much on the learner's first language, etc. As a result, Grammar-translation was almost banished from the scene of language education at the latter half of the twentieth century. Nevertheless, the grammar part of it has always stayed with the profession in some form, though the translation part was largely abandoned. In recent years, however, as Cook (1998) observes, there is a revival of interest in including translation as one of the activities in the language classroom. Cook correctly points out the value of translation by saying that Grammar-translation "holds no monopoly" in using translation for second language instruction, and that translation can actually be used "more imaginatively" in a language classroom.

2. The Direct method

This method became popular after Grammar-translation lost its public favour. Its characteristics are exactly opposite to those of Grammar-translation. It bans the use of mother tongue in class and thus translation is not allowed. Classes are conversation-based and the teacher must be a native or near-native speaker of the target language (Celce-Murcia, 1991). The method works well in private language schools with highly-motivated paying clients and the employment of native-speaking teachers. However, it suffers a lack in theoretical foundations, relying heavily on the teacher's personal skills in implementation. It emphasises naturalistic language acquisition scenarios too much, relying too much on speech at the expense of forms, and fails to consider formal classroom settings. Thus the method is scarcely applicable in public school settings (Richards & Rodgers, 1986).

3. The Audiolingual method

The Audiolingual method emerged towards the end of the 1950s in the United States as the U.S. Government's response to the needs to keep being informed of the

scientific advances in foreign countries. The theory of the method was derived from the structural linguistics popular at that time, which emphasises the collection and analysis of spoken data according to a structured system of phonemes, morphemes, words and sentence types (Richards & Rodgers, 1986). Another important theory at the time which contributed to the learning theory of the Audiolingual method was behavioural psychology, which claimed animal (including human) learning behaviours to be a kind of 'habit formation' consisting of a stimulus, a response, and a reinforcement. Thus Celce-Murcia (1991) says, in audiolingualism, "Mimicry and memorization are used, based on the assumption that language is habit formation". The notorious 'pattern practice' in audiolingualism, according to Stern (1992), is still valuable today to some extent in that "it provides an opportunity for exercising language patterns in a way which is cognitively undemanding" (p. 341). The method still has its value as long as it is not used in isolation.

As for the attitude towards translation, because Audiolingual considers speech to be primary and writing to be secondary, it actually does not offer learners systematic training in reading and writing. Instead, "students ordinarily learn something of the processes [reading and writing] as a by-product or in support of oral-aural language study" (Croft, 1972). Translation, therefore, is secondary at best to the audiolingualism since it is a reading-writing activity as opposed to interpretation. Thus, according to Richards & Rodgers (1986: 58), quoting Brooks (1964), one of the procedures for a teacher in Audiolingual to follow is: "Practice in translation only as a literary exercise at an advanced level".

4. COMMUNICATIVE LANGUAGE TEACHING

According to Richards & Rodgers (1986:64), Communicative Language Teaching (CLT) originated in Britain in the late 1960s. For CLT, the goal of language teaching is for students to have communicative ability; thus classroom activities often consist of students working in pairs or groups and making meaning-based interactions (Celce-Murcia, 1991). To learn communicative skills in the second language and to be able to get meaning across is the utmost concern and grammar and error corrections are secondary.

Howatt (1984) concludes the discussion on CLT by saying, "The overall impact of the communicative approach has been to enrich and extend the traditions of language teaching initiated by the reformers at the end of the last century" (p. 289). And the 'traditions' referred to by Howatt are: the emphasis on the spoken language, the connected discourse, and the monolingual principles, etc. (ibid.). The emphases on spoken language and the monolingual principle indicate that translation will not be

a favoured teaching activity in CLT. According to Richards & Rodgers (1986: 67) citing Finocchiaro & Brumfit's (1983) contrasts between Audiolingual and CLT, for Audiolingual, translation "is forbidden at early levels"; whereas for CLT, translation "may be used where students need or benefit from it". In sum, though translation is not explicitly and entirely banned in CLT, it certainly is not encouraged.

5. AFFECTIVE-HUMANISTIC APPROACH

This is an umbrella term for innovative approaches like Community Language Learning (CLL) and Suggestopedia, both making regular use of translation. In Suggestopedia's unique 'music listening and reading' session, the teacher reads the text for learning and the students look at the text which was prepared both in the target language and in the native language (Richards & Rodgers, 1986: 151). In the case of CLL, Richards & Rodgers (1986: 120) describes how translation is used:

> Learners form a small circle. A learner whispers a message or meaning he or she wants to express, the teacher translates it into (and may interpret it in) the target language, and the learner repeats the teacher's translation.

In both cases, translation is used as a bridge to pave the way for the learner to cross from his native language to the target language. In these humanistic approaches, resorting to the mother tongue is one of the ways of filtering out the learner's anxieties. In other words, translation is in the service of the affective domain rather than carrying the burden of some instructional purposes.

3.2.2 SLA theories

The role of translation in second language acquisition (SLA) has not been extensively discussed by researchers in this field. The literature is abundant however, in a related thread of discussion – the problem of language transfer in SLA. Although some researchers doing contrastive analysis (CA) claim in the beginning that similarities between two languages will facilitate transfer from L1 to L2 (positive transfer); while dissimilarities will cause problems in learning the L2 (negative transfer), the position in its strong form has largely been abandoned since, as Ellis (1994) notes, "many errors are not caused by transfer" and "many predicted errors do not occur" (p. 342). The contrastive analysis hypothesis is not entirely wrong, however, in that it recognises the importance of the role L1 plays in L2 acquisition, as Ellis concludes in his discussion on language transfer: "there is now clear evidence that the L1 acts as a major factor in L2 acquisition" (p. 343).

The exact role L1 plays in SLA, however, is very difficult to determine. Significant though it may be, it is difficult to decide how L1 should be made use of in the second language classroom. Pica (1994), for example, asks: "In what ways does knowing one language help or hinder the learning of a second?" in addressing the problems associated with a language classroom from the research perspectives. While also recognising that "predictions of contrastive analysis do not always hold" in that "L1 and L2 differences do not necessarily imply learner difficulties, and similarities between L1 and L2 features do not guarantee that all of them will be learned with ease", the best conclusion she can arrive at in terms of the use of L1 in the L2 classroom seems to be that "the influence of the learner's L1 is highly selective, far more intricate, and less readily predictable than was previously thought".

Some positive answers are given by Auerbach (1993) as to how L1 can be helpful to learners in the L2 classroom. Auerbach's discussion starts from her observation that "when the native language *is* used, practitioners, researchers, and learners consistently report positive results" (p.18). One of the ways L1 can help L2 learners is in writing, where students first write in L1 and then translate into L2. Auerbach refers to Osburne & Harss-Covaleski's (1991) research, which suggests that "the widely frowned upon practice of writing first in the L1 and then translating into the L2 is not detrimental to the quality of the written product" (p.21). The theoretical foundation for this phenomenon may be that "strong initial literacy is a key factor in successful second language acquisition and academic success" (p.15), which coincides with Ellis' (1994: 339) model of language transfer, where L2 learners' interlanguage is generated from the L1 system through the processes of "borrowing" and hypothesis construction and subjected to some sorts of constraints.

Another thread of research in SLA is the "consciousness" (or implicit vs. explicit learning) debate (Ellis, 1994: 359-362) which can also be related to the practice of translation in L2 learning. A well-cited work in this connection is Schmidt (1990), who claims "that subliminal language learning is impossible, and that noticing is the necessary and sufficient condition for converting input to intake". That is, noticing (which can be achieved by, for example, task demands or word frequency) must take place if the learner is to learn the second language. In the L2 teaching battlefield, so to speak, Tarvin & Al-Arishi (1991) would like to re-examine CLT (communicative language teaching) in terms of the element of "reflection" in EFL. They suggest that "CLT with its emphases on conspicuous action and spontaneous response has unintentionally slighted the need and desire of language learners to abstract, generalize, and synthesize". Arguably, the learner's need to "abstract, generalize, and synthesize" could coincide with the "noticing" effect proposed by Schmidt. Again arguably, one of the most effective means for learners to notice and reflect on

the language would be for them to do translation between L1 and L2. Cook (1998) summarises for us in saying that there is a growing awareness that "translation can, as it was traditionally believed to do, develop accuracy", since "translation can focus attention upon subtle differences between L1 and L2", among other things.

3.3 Translation teaching

In contrast to second language teaching, the teaching of translation does not have a large body of research behind it. Newmark (1991), at the beginning of his chapter *Teaching Translation*, says, "I approach my subject, teaching translation, rather apprehensively. To my knowledge, there is hardly any literature on it" (p. 129). However, Newmark considers translation teaching a promising new profession and wishes to distinguish it from "teaching translation within language teaching" (p. 137), in which, as observed by Bussmann (1996), translation is a 'fifth skill' (next to speaking, listening, reading and writing) merely "used to practice and test competence and performance in a second language".

With respect to the meagreness of historical asset in theories of translation teaching, Caminade & Pym (1998:280) also point out:

> Translators and interpreters have long been trained informally, basically through trial and error, unstructured apprenticeship arrangements, or any of the various translating activities that accompany the study of a foreign language and culture within the Liberal Arts tradition.

In fact, the backwardness of the translation education in the United States leads as recent a translation practitioner as Sofer (1998) to say that "most working translators in America today were trained, so to speak, on the job" (p. 145).

Meagre as the literature may be, Vermeer (1998), nevertheless, is able to discern two types of translation teaching -- the traditional approach and the functional approach. The following describes how translation is traditionally taught:

> Translational skill is acquired by exercises based on linguistic equivalence rules of the type: 'translate German adverbs by a Spanish final verb + *que* construction and vice versa ...'.

However, Vermeer claims, the teaching of translation is no longer the same with the advent and elaboration of the functional approach to translation. The functional approach advocates that a text is produced in a specific context with specific recipients, communicative functions, and so on. So the translation of a text cannot

be limited to the transformation of the text from one language to another alone. It must involve the handling of the source and target communicative contexts. In other words, the translator should be equipped with not only linguistic competence but also cultural competence. This point is amply endorsed by recent researchers in translation and translation teaching. For example, Krouglov (1996) says:

> A deep understanding of the ways in which social and cultural features are combined in a language is indispensable in teaching translation and interpreting.

Also, Ulrych (1996:253), while advocating the integration of vocational and educational components in teaching translation, also claims:

> In order to deal appropriately with a text, translators need to know all the characteristics of the text they are translating ... as well as its status in the source culture and the intended audience in the target culture.

In sum, it is perhaps safe to say that although translation pedagogy has been unfairly neglected in the past, it is now catching up with the progress in translation studies and the continuing demands for translation and high-quality translators. The view of translators as communicators between two or more cultures is now widely accepted and informing the didactics of translation accordingly. A point to note is that, in our model of translation competence for second language learners (expressed in Figure 3-1), this ability to mediate communication between cultures is correctly subsumed in our sociocultural component of competence.

3.4 An integrated model

At the end of the discussion on the competence and pedagogy of language and translation, I would like to posit a model which reflects the essence of this thesis, i.e. for second language learners learning to translate into the second language to acquire translational skills and proficiency in the second language at the same time. This is different from the "teaching translation within language teaching" practice discussed above, as our main focus is on teaching translation rather than language learning. This also differs from translation teaching in the usual sense since we are dealing with translators as target language learners, rather than translators fluent in both source language *and* target language. The entire concept is roughly represented in Figure 3-2.

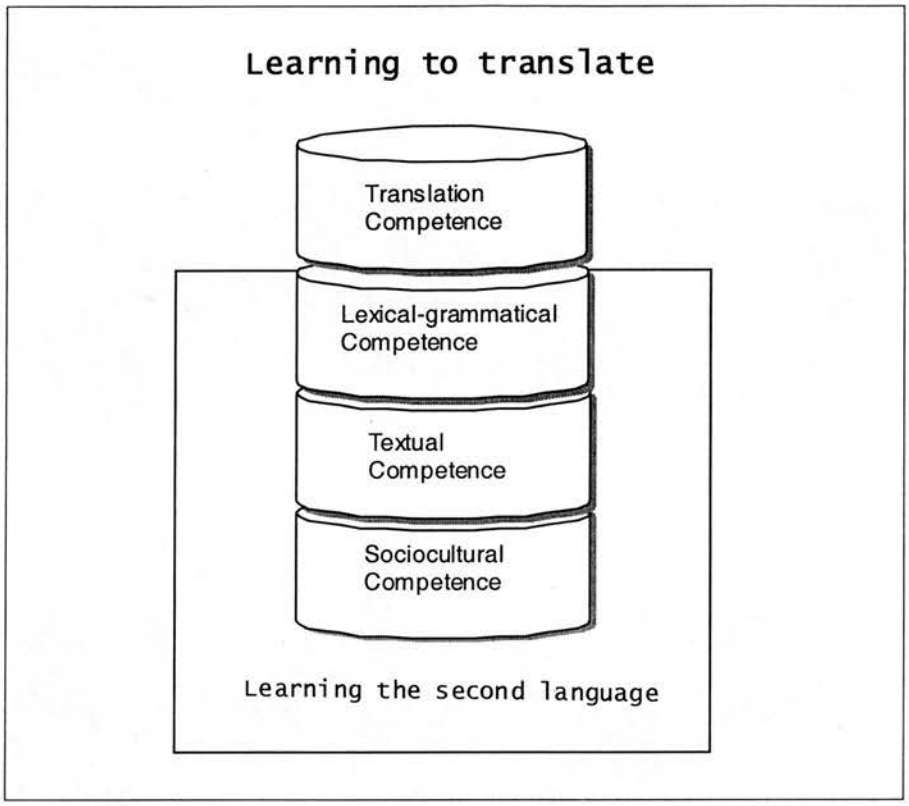


Figure 3-2: Learning to Translate into the Second Language

Figure 3-2 depicts a model of the learner who is learning to translate and learning the target language simultaneously. In this model, learning translation subsumes the learning of the second language. The figure shows translation competence to be built upon second language competence consisting of three components -- lexical-grammatical, textual and sociocultural competence. Hence the student's ability to translate into the second language cannot develop satisfactorily without his competence in the target language being improved on first. The model makes it clear that any translation pedagogy for the circumstance prescribed must take the language learning factors into consideration as well. Any teaching methodology proposed to deal with the education of translators who are learners in the target language should endeavour to conform to this principle. In Chapter Five, an instruction model will be proposed on the basis of the model depicted in Figure 3-2.

3.5 Summary

This chapter endeavours to establish a model of translation competence with special reference to translating into a second language. It starts by examining the distinction between competence and performance, recognising that the two aspects of language ability actually build on and help refine each other. Three aspects of linguistic

competence are then isolated based on various models proposed by researchers: lexical-grammatical competence, textual competence and sociocultural competence. On top of this core, translation competence comprising monitoring competence, hi-tech competence, content-knowledge competence, etc. is added. It is argued that, for a translator translating into L2 to be successful, the linguistic basis must be reasonably accomplished before translation competence can become useful. This chapter also describes some mainstream L2 teaching methodologies in relation to the use of translation. This provides a reference point for the translation teaching pedagogy to be proposed in later chapters.

Chapter 4

Learner Corpora and Error Analysis

This chapter presents two analyses of learner corpora – a composition corpus and a translation corpus respectively. The analyses centre around the collocational and idiomatic abilities of the learners as observed from the data, and their relation to the learners' ability in grammar. The purpose for doing so is to uncover some of the learners' weaknesses in their interlanguage, which is valuable information for designing instructional goals, methods, means, materials and activities.

A learner corpus is useful in many ways, one of which is to diagnose learners' errors so as to render pedagogical solutions. In describing The International Corpus of Learner English (ICLE), Granger (1996) claims that the main objectives of the project are "to uncover the factors of non-nativeness or foreign-soundingness in advanced learner writing" and "to distinguish between L1-dependent features ... and crosslinguistic invariants". The first objective reflects a well-noted point that advanced learners of English as a Foreign Language generally have fewer problems in grammar than in the choice of words in their writing, including collocations (Shei & Pain 2000), so-called lexical phrases (Nattinger & DeCarrico 1992) and lexicalised sentence stems (Pawley & Syder 1983). The second objective attempts to pin down whether the reasons for the unnaturalness are due to mother tongue influence or a universal inclination to make these errors. The first objective is directly relevant to this thesis, since its purpose is to depict a computerised language learning environment in which learners can enhance their abilities in writing English with respect to 'naturalness' or word selections. To show how word selection abilities can be measured, this chapter discusses a machine-aided way of analysing collocations in learners' writings. To investigate how the ability to write idiomatic phrases is related to the ability to write grammatical sentences, this chapter also offers an innovative way of analysing a learner translation corpus.

Chang Jung University (CJU) is a small university located in the south of Taiwan. The department of translation, where the current author works as an associate professor, consists of 566 students as of the academic year 2000. English is the primary foreign language being studied, with other secondary foreign languages being German, French, and Japanese. Translation from and into the foreign languages are equally emphasised. Also, not only translation modules are offered, but language enhancement modules in the foreign languages such as grammar, composition and

conversation are also obligatory.[10] Since October 2000 the English composition and English-to-Chinese and Chinese-to English translation exercises and examination papers of the Department's students taught by this author have been collected to form a composition corpus and a translation corpus: both are learner corpora.

4.1 The composition corpus

The sample writing data analysed in this section are composed of one assignment of one of the author's several English composition classes taken randomly from its mother corpus and consisting of 29 short compositions amounting to approximately ten thousand words. The analysis focuses on the collocational aspects -- for both lexical and grammatical collocations -- rather than on the more creative side of grammar or semantics, as collocations are relatively more fixed and are thus good candidates for computer modelling for the purposes of either research or construction of intelligent tutoring systems.

4.1.1 Collocation analysis

In contrast to Howarth (1998) who conducted manual inspection on his 25000-word learner data to extract 'verb + noun' collocations, this author uses a machine-aided method which extracts verb + noun candidates from the corpus for quick inspection and identification of deviant collocations.[11] It is hoped that the initial thoughts presented here may inspire more research on the computational modelling of learners' collocational abilities which is of crucial importance to the design of CALL software dealing with collocations.

First the entire corpus should be run through a part-of-speech tagger, which is a computer program designed to find the part of speech for every word in a given file.[12] Below is an example that contrasts tagged with untagged texts for a line taken from the sample corpus:

---

[10] See Chapter 11 for a more detailed discussion on the courses of this department.

[11] By "deviant collocations" I mean collocations which are not normally used by native speakers. Objective judgment of these can be derived from corpus evidence, as will be explained later, especially in 5.3.2.

[12] Note tagging is a serious topic in computational linguistics and is not explored farther in this thesis. Note, however, that automatic tagging for native speaker texts is not 100% correct, not to mention that for learner input which contains frequent errors. The result of automatic machine tagging for learner corpora thus has to be edited in some way by human hands, or separate tagging algorithms must be developed for dealing with learner corpora.

(1)

a. Those would paint untrue imaginations for youngsters and directly affect youngsters' sexual conceptions.

b. Those_DT would_MD paint_VB untrue_JJ imaginations_NNS for_IN youngsters_NNS and_CC directly_RB affect_VB youngsters_NNS '_POS sexual_JJ conceptions_NNS ._.

Another computer program then takes over the task of extracting the verb-noun pairs from the tagged corpus, each verb tagged with VBX being the 'node' word with a scope covering three words to the right of the node, within which range any word marked with the NNX tag will be extracted.[13] There are altogether 586 types of verb-noun pairs extracted from the 10000-word corpus, an excerpt of which is shown below:

(2)
```
made breathing 1
made laws 1
made short 1
made surprise 1
make can 1
make crime 1
make direction 1
make impression 1
make influence 1
make law 1
make laws 2
make love 1
make matter 1
make money 1
make overstatements 1
make people 1
make teenagers 1
make things 1
makes guys 1
makes scope 1
making sick 1
mans skills 1
mans watch 1
misguide youngers [sic] 1
mislead someone 1
misleads strippers 1
mix idea 1
```

The number following each verb-noun pair in (2) is the frequency of the pair in the sample corpus. However, in this qualitative analysis, we are only concerned with what types of collocational errors there are, rather than counting the frequencies of each type. Thus the next step is to manually inspect the 586 lines of verb-noun pairs like the above and try to pick out the deviant ones – those not seemingly conforming

[13] The scope of three words, i.e. Verb+Article+Adjective+Noun, should cover most situations where a verb is followed by an object noun.

to the native speaker's choice.[14] In doing this small-scale analysis, the author, being an advanced learner of English as a foreign language, consulted his own intuition as a first step of judging the legitimacy of the collocations. For any collocation whose status the author was unsure of, to verify its legitimacy or abnormity, the next step was to consult a corpus, namely the British National Corpus or the Bank of English, to see whether native speakers have indeed used the collocation. If native speakers used the same collocation in identical or similar circumstance, then it was passed as good. If no example was found in the corpora consulted, then the Google Web search engine was used to try to locate related collocations from native speakers' writings.[15] If the Web search did not yield positive results, then the suspicious collocation was assumed to be an abnormal usage.

So for example in (2), the pairs *made breathing*, *made surprise, make crime, make direction* and *make influence*, are picked out as likely to be unusual collocations. They are first verified as truly being used by the students as verb-noun collocations by the students by being presented in their sentence context, as they appear in the corpus. Thus for this example the following sentences are picked out and scrutinised:

(3)
a. My mother **made** a deep **breathing** and said, "you shouldn't chat with strangers because it's too dangerous...."
b. At first, when I received the pornography, it **made** me very **surprise**,...
c. And they usually use Internet to **make** a **crime**.
d. Besides, pornography will **make** incorrect **direction** on sexuality for juveniles.
e. Besides, her thought is the same as me that the pornography will **make** bad **influence** on the children.

It can be seen that the student writer does not mean for *made surprise* in (3b) to be a verb-noun collocation but part of a verb-object-compliment construction instead. The rest of the examples in (3) do show badly improvised collocations involving the two forms of *make*. This is partially verified by consulting the COBUILD or BNC corpus and finding no examples of use of these collocations.[16]

To evaluate the effectiveness of this collocation extraction procedure, a subsequent inspection is performed on the sample corpus on paper to manually pick out all the irregular verb-noun collocations. In judging the appropriateness of the

---

[14] There is also a computational means for automatically spotting suspicious collocations, which will be mentioned in later discussions.

[15] The use of Google to look up collocations is more tricky since its design is not facilitative to this kind of research. However, it is still very useful and the author regularly uses this to supplement the inadequacy of corpus evidence in teaching.

[16] Although there are examples of *make direction* in COBUILD, the word *direction* is used in the sense of "instruction", not that of "path pointing" as the student intends.

verb-noun collocations, the CobuildDirect Collocation Sampler (CDCS) was used.[17] If there turned out to be one or more instances of the suspicious verb-noun combination in the CDCS corpus, the pair was passed as OK. Otherwise it is marked as a deviant collocation. The results are compared with those made by the machine-aided method, as Table 4-1 shows:

Deviant verb-noun collocations found by human	**Possible source of error**	**Also found by machine or not**	**Reason why machine failed**
absorb articles	L1 influence	yes	
affect value	L1 influence	no	noun intervening
build thoughts	L1 influence	yes	
cause ideology	L1 influence	yes	
cause influence	L1 influence	no	verb tagged as noun
communicate materials	L1 influence	yes	
contact equipments	L1 influence	yes	
damage aspects	L1 influence	yes	
damage youngsters	L1 influence	yes	
did attacks	general verb effect	yes	
eat medicine	L1 influence	no	noun intervening
experience lesson	L1 influence	yes	
find laws	L1 influence	yes	
get conceptions	L1 influence	yes	
have spirits	L1 influence	yes	
lowers years	L1 influence	no	noun intervening
made breathing	general verb effect	yes	
made guiltiness	general verb effect	no	inverse construction
make articles	general verb effect	no	verb tagged as noun
make crime	general verb effect	yes	
make direction	general verb effect	yes	
make influence	general verb effect	yes	
notice responsibilities	L1 influence	yes	
obstruct lives	L1 influence	yes	

[17] This is on http://titania.cobuild.collins.co.uk/form.html at the time of writing. The program automatically searches a large English native-speaker corpus and finds collocates for the query word, listing them in the order of statistical significance.

offers ideas	L1 influence	yes	
paint imaginations	L1 influence	yes	
prevent security	misunderstanding of word meaning	yes	
read pornography	L1 influence	yes	
research information	misunderstanding of word meaning	yes	
shock mentality	L1 influence	no	inverse construction
solve code	L1 influence	yes	
solve codes	L1 influence	yes	
stress development	L1 influence	no	verb tagged as noun
teach knowledge	L1 influence	yes	
tell conception	L1 influence	yes	
watch E-mail	L1 influence	no	noun intervened
watch pornograph	L1 influence	yes	
watch pornography	L1 influence	yes	
watched pornography	L1 influence	yes	
watching pornography	L1 influence	yes	

Table 4-1: Deviant learner verb-noun collocations

There are actually two different kinds of information encoded in Table 4-1. First, how many kinds of deviant verb-noun collocations have the students produced in the sample corpus and what are the possible sources of errors? Second, how many pairs of these collocations has the machine-aided method failed to retrieve and why? We discuss the second issue first.

As Table 4-1 shows, manual inspection of the sample writing corpus has yielded 40 instances of deviant verb-noun collocations, of which 9 were not discovered by the machine-aided method. There are three kinds of reasons to account for this failure: because there is another noun intervening between the verb and the true object of the verb, because the verb itself is tagged as a noun by the tagger and so is ignored by the verb-noun extraction program, and because the pair appears in an inverse construction in which the position of the verb and that of the noun are reversed. Each type of retrieval error is illustrated by an example below:

(4)

a. Intervening noun:

her_PRP$ husband_NN **eats**_VBZ a_DT kind_NN of_IN **medicine**_NN ,_,

b. Verb tagged as noun:
Because_IN pornography_NN **cause**_NN harmful_JJ **influence**_NN on_IN youngsters_NNS ,_,

c. Inverse construction:
their_PRP$ **mentality**_NN will_MD be_VB **shocked**_VBN seriously_RB ._. (passive)
he_PRP still_RB didn't_VBD know_VB what_WDT **guiltiness**_NN he_PRP **made**_VBD ._. (relative clause)

In (4a), the deviant collocation *eats medicine* was not spotted by the computer program because there was another noun intervening, namely *kind.* Thus *eats kind* was retrieved but not *eats medicine.* The problem could possibly be corrected by increasing the span to four words when encountering quantifier phrases such as *a kind of, two pounds of* etc. in order to fetch the relevant object noun beyond. That is, when the algorithm encounters examples like *drink an ounce of milk*, it will ignore *ounce* and increase the span to get the next noun, i.e. *milk.* However, expanding the window also increases the system's burden of processing and the chance of getting more irrelevant V-N pairs. In (4b) the verb *cause* was incorrectly tagged as a noun, so that the verb-noun collocation *cause influence* was missed out altogether. This is a problem residing in the tagger program and is outside the scope of this thesis.[18] In (4c), the original order of verb-noun is reversed in both clauses so that the computer program specifically written for that order failed to discover the deviant collocations. More computer codes could conceivably be written to enhance the processing power of the program in this respect, by incorporating rules specifying conditions for the inverse of the verb and the noun such as in a passive or relative clause construction.

Altogether, the performance of the machine-aided collocation retrieving method can be summarised in terms of **precision** and **recall**, as defined in Oakes (1998: 176):

$$precision = \frac{Number\ of\ good\ candidates}{Total\ number\ of\ candidates}$$

$$recall = \frac{Number\ of\ good\ candidates}{Number\ of\ terms\ in\ the\ reference\ list}$$

Thus precision is the proportion of the good candidates found among all the candidates being retrieved. Recall is the number of good candidates found against

[18] See Garside et al. (1997) for a useful discussion of tagging.

the number of the qualified terms there are in the database. In our case,

$$precision = \frac{31}{586} = 0.0529$$

$$recall = \frac{31}{40} = 0.775$$

Thus the machine-aided collocation retrieval method in its current form has acceptable recall but very poor precision measure, which means that the program is reasonably good at retrieving the desirable records (or that the program leaves very few false negatives behind, which are good candidates that are passed as unqualified), but it also retrieves lots of undesirable ones; that is, it retrieves a large number of false positives, or pairs which seem to be good candidates when they are not. This is because of the huge list of normal collocations retrieved as compared to the small list of irregular ones. In analysing learner corpora in terms of collocations, as each instance of deviant collocation is a valuable object for study, it may be more fruitful to take a liberal approach as opposed to a conservative one in order not to miss any good candidates.[19] Shei & Pain (2000) introduce the concepts of **collocation library** and **error library** which can drastically decrease the number of incorrectly selected normal collocations extracted and increase the precision rate. Now we move on to discuss the nature of the collocational errors in the students' data.[20]

The verb-noun collocation is not only frequently the first kind of collocation to be mentioned or investigated (Benson et al. 1997, Howarth 1998, Smadja 1993), it is also found to be the most difficult type of collocation for ESL learners (Gitsaki 1996). Recall that one of the main objectives of the International Corpus of Learner English is to find whether the unnaturalness in learner writings is due to L1 influence or a universal phenomenon. In the case of CJU students' writing, Table 4-1 shows that most deviant verb-noun collocations produced (31 out of 40) could be the results of mother tongue influence, since they are all direct (i.e. literal) translations of their corresponding original Chinese phrases, which happen to be bad collocations in English. One obvious example is *eat medicine* (which is clearly an unacceptable deviant of *take medicine*). In Chinese to transfer medicine into one's stomach is literally said to be 'eating medicine' (*chi-yao*), just like the way one 'eats cooked rice'

[19] The situation is similar to that of an intrusion detection system, where the most important thing is not to create any false negative error, which means an act of intrusion is judged by the security system as harmless. See http://www.cerias.purdue.edu/coast/intrusion-detection/detection.html.

[20] Note that the N-V type of collocation, with N representing the subject and V the main verb of a sentence, is also a meaning kind of collocation, which however is not dealt with here because the automatic extraction of such is beyond the scope of the simple computational method used here.

(*chi-fan*), hence the negative lexical transfer to English. The other example is *watch E-mail* or *watch pornography*. In Chinese the character corresponding to *watch* here is *kan*, which denotes the meaning of 'capturing information through the eyes', no matter what forms the information is in. Thus *kan-shu* means 'read books', *kan-wo* means 'look at me', *kan-dianshi* means 'watch television', and *kan-fengjing* means 'appreciate the view'. Somehow many Chinese students internalise the rule that *kan* corresponds to the English word *watch*, which could only account for the fact that Chinese learners frequently produce unnatural verb-noun collocations like *watch newspapers*, *watch the flowers* (meaning 'appreciate the flowers') etc.

The second frequent reason for CJU students producing deviant verb-noun collocations is what might be called the 'general verb effect' in Table 4-1. This happens when the student has not internalised the specific verb to describe an action and resorts to using general verbs like *make*, *take* and *do*, etc. This can be surmised from the fact that these kinds of verbs are often seen in learner corpora in place of more specific verbs for a certain state or action. For example, Table 4-1 shows that the student using *make crime* does not know *crime* is habitually collocated with the verb *commit* in English, so she uses *make a crime* instead, assuming that in English *make* is an all-around verb suitable in many kinds of context. Finally, for the other two deviant forms of verb-noun collocations in Table 4-1 which are not accountable by the above two main factors, their original contexts are printed below to allow further consideration:

(5)

a. I also will watch this pornography in several websites while I **research** some **information** for my translation homework.[21]
b. In order to **prevent** the **security** in our life, we should make some laws to control the pornography.

Clearly in (5a) the student confuses *research* with *search (for)*, which she really means to say. In (5b) the student most likely has used *prevent* to mean something like 'preserve'. In both cases the learner misunderstands the meaning of a word (the verb), which is not a typical collocational error. The meaning of a typical unnatural collocation is usually retrievable, but the form may be irritating in some way to native speakers. Collocation is a main pedagogical focus of this thesis, and we will return to it in Chapter Five.

---

[21] As their teacher, the author is quite positive that the student really means "I also will be forced to view pornographic materials accidentally when I search for information, which just automatically pop up as advertisement from some innocent-looking sites". This may be a good example of what a bad learner collocation can mislead the reader to think.

4.1.2 Grammar analysis

Unfortunately current parsing technology does not allow efficient large-scale error analysis of corpora beyond the levels of words and phrases. Some CALL programs have attempted to automatically detect grammatical errors in learners' writing, but these tend to work only in particular aspects of grammar or for certain forms of input. For example, Schuster's (1996) 'VP^2'system deals with the checking and correcting of ESL learners' verb plus particle or verb plus prepositional phrase constructions; Wang & Garigliano (1992) tackles only the problems of English-speaking students learning Chinese caused by transfer; Bull et al. (1993) focus on student modelling in the learning of certain morphological rules in Portuguese. Sentence (1997) introduces an ITS which allows free text input but which deals with the use of articles only in English. Finally, Bolt (1993) compares the performance of six commercial grammar checking programs with LINGER -- a system developed at Exeter University. Bolt claims that the commercial programs all fail to pick out the errors in some way because they "rely on various forms of pattern matching techniques" only and do not "capture the complexity of the data". The academic system, LINGER, on the other hand, "has the virtue of dealing with a small part of the data reasonably well" but it cannot "continue to perform as well with larger data sets". That the commercial systems still perform poorly on grammar checking to date can be seen in the following text excerpted from the sample writing corpus, which was run through the grammar checker of Microsoft Office 20000 SR-1.

> (6)
> Most teenagers are curious about the sex, but they're no proper way to get the right knowledge about sex; the easily-get information comes from the Internet. The users just type the relative words, such as "sex", "pornography", and search it, then, soon, they will get hundred of information about pornography. The pornography offers the wrong ideas that make immature teenagers think it is true. Also, School teachers and parents don't know how to avoid these from their children.

A casual reading of the passage would have revealed several problems in it, grammatical or lexical. Table 4-2 summarises just some of the problems the author, as a non-native speaker of English, can identify:

Error	Correction	Description
the sex	sex	article redundancy

they're	there are[22]	similar words confusion
way	ways	plural form omission
easily-get	easily-found	passive form confusion, lexical choice
relative	related	confusion of derivational forms
search it	search for them	pronoun-antecedent number mismatch, verb transitivity confusion
hundred of information	lots of information	quantifier-noun mismatch
The pornography	Pornography	article redundancy
the wrong ideas	wrong ideas	article redundancy
avoid these from	keep these from	wrong verb subcategorisation

Table 4-2: A summary of errors for student writing excerpt in (6)

However, the Office 2000 grammar checker neglected all these problems and complained about only one "error" -- that *School teacher* should be changed to *Schoolteacher*, which the author did not see as a problem at all.[23]

The above discussion amounts to saying that large-scale grammar checking seems currently unreachable, and a good way to do grammatical error analysis is first running the text through machine and then examining it by hand. However, as we shall argue in the next chapter, grammar may not be the most important aspect of language production: prefabricated items may be equally, if not more, important in the processing of language.

For this and other reasons, it may be more fruitful to look at the so-called **grammatical collocations** in this section, which is defined in Benson et al. (1997) as "a phrase consisting of a dominant word (noun, adjective, verb) and a preposition or grammatical structure such as an infinitive or clause" (p. xv), as opposed to **lexical collocations** which "normally do not contain prepositions, infinitives, or clauses" and "consist of nouns, adjectives, verbs, and adverbs" (p. xxx). By this definition, the verb-noun collocations appearing in the previous section are all lexical collocations, so are adjective-noun collocations like *strong tea* and *rough estimates* and adverb-adjective collocations like *readily available* and *deeply absorbed*; while

[22] Helen Pain points out that this should be either *there is* or *they have*, but from the author's point of view both as an advanced learner and as an EFL teacher to Chinese students, *they're* is more likely the mistake of *there are* because of their strong phonetic similarity. In other words, the learner might have internalised the pronunciation of *there are* and its meaning, but has realised it in a different orthographic form, i.e. *they're*. Since the learner does not have very strong grammatical instinct, the phonetic influence overrides the grammatical concerns.

[23] This performance does not represent state-of-the-art technology or research in grammar checking, of course. It is simply a casual inspection of what a representative commercial word processor can do in terms of grammar checking.

grammatical collocations are phrases like *account for* and *apathy towards* (a content word plus a particle) and structures like *offer to help*, *a pleasure to do something*, and *afraid that it would rain* (a content word plus an infinitive phrase or a clause). Since grammatical collocations cover plenty of the traditional ground of grammar in language teaching, and can be dealt with more easily and efficiently in a computer environment, in this thesis more emphasis is laid on grammatical collocations than on the more creative or productive grammatical rules when dealing with grammar both in error analysis and the design of intelligent tutoring.

That grammatical collocations can account for a considerable proportion of grammatical errors in learners' writing can be illustrated with an example, a sentence taken from the sample corpus:

(7)
Most parents who have younger children **worry about that** they could not do anything to **avoid their children looking** the pornography on the net.

One thing that is striking about the sentence in (7) is its quite sophisticated clausal structure in contrast to the inherent awkwardness: It has a sound relative clause (*who have younger children*) and a basically correct noun clause stem (*that they could not do anything* ...). However, the sentence still does not read naturally, the reason being that the sentence plays host to three grammatical collocation errors. Namely, the student confuses the *worry about* (verb + preposition) collocation with the *worry that* (verb + clause) collocation and produces an erroneous blend. Secondly, the student incorrectly places *avoid* in the 'verb + noun + gerund' (such as *keep the car moving*) grammatical collocation category and produces a deviant structure. Finally, the student has not internalised the grammatical collocation *look at* (verb + preposition) and mistakes *look* for a transitive verb. If we correct these three grammatical collocations, we get a relatively well-formed English sentence:

(8)
Most parents who have younger children **worry that** they could not do anything to **prevent** their children **from looking at** the pornography on the net.

Because grammatical collocations are less creative and hence computationally more tractable, it is sensible to deal with these first both in automatic error detection and in computerised writing instruction, until a full-fledged parser can be invented and utilised in a CALL environment. In this connection, a computer program can conceivably be written which can extract the eight types of grammatical collocations distinguished in Benson et al. (1997), based on part-of-speech information and working on POS tagged texts without having to rely on sophisticated parsing

technology. Table 4-3 below shows all the grammatical collocation errors manually extracted from the sample corpus.

Incorrect grammatical collocation	**Correct usage for intended meaning**	**Correct grammatical collocation category**
advise you don't	advise you not to	verb + NP + infinitive
appeal youngsters	appeal to youngsters	verb + Prep
approaching to pornography	approaching pornography	(none)
avoid children touch	prevent children from touching	verb + NP + verb-ing
avoid their children getting	prevent their children from getting	verb + NP + verb-ing
avoid their children looking	prevent their children from looking	verb + NP + verb-ing
being contact with	being in contact with	Prep + noun + Prep
can access to the pornography	can gain access to the pornography	noun + Prep
cause youths deviating	cause youths to deviate	verb + NP + infinitive
contact with sex	contact sex	(none)
control their children to the pornography	control their children's access to pornography	(none)
curious what the pornography is	curious about what pornography is	adjective + Prep
happen on	happened to	verb + Prep
influence for	influence on	noun + Prep
insisted to send	insisted on sending	verb + Prep
is harm for	is harmful to	(none)
keep touch with	keep in touch with	Prep + noun + Prep
lead them to wrong directions	lead them in wrong directions	Prep + noun
limit people to access	limit people's access to	(none)
listening my classmates' explain	listening to my classmates' explain	verb + Prep
look pornographic and women	look at women and use pornography	verb + Prep
make our society can purified	make our society purified	verb + NP + adjective
make them can't pay	make them unable to pay	verb + NP + adjective

makes those bad guys can do it	makes those bad guys able to do it	verb + NP + adjective
makes us to worry	makes us worry	verb + NP + verb-root
must to prevent	must prevent	auxiliary + verb
out of curious	out of curiosity	Prep + noun
pay more attention on	pay more attention to	verb + noun + Prep
prevent Internet pornography to affect youngsters	prevent Internet pornography from affecting youngsters	verb + NP + V-ing
prevent teenagers from pornography	keep teenagers from pornography	verb + NP + Prep
prevent the kids from the damage	keep the kids from the damage	verb + NP + Prep
prevent youths entering	keep youths from entering	verb + NP + V-ing
proud by	proud of	adjective + Prep
provide us a clean environment	provide us with a clean environment	verb + NP + Prep
respect for others	respect others	(none)
saw a black naked man appealed	saw a black naked man appear	verb + NP + V-ing
see the pornography appears	see the pornography appear	verb + NP + V-ing
stop publish	stop publishing	verb + V-ing
suggest that my aunt to buy	suggest that my aunt buy	verb + NP + infinitive
take care about their children	take care of their children	verb + noun + Prep
take notice on	take notice of	verb + noun + Prep
tell a wrong conception toward youngsters	give the wrong ideas to youngsters	verb + noun + noun
to access to the Internet	to gain access to the Internet	noun + Prep
worry about that	worry that	verb + noun clause

Table 4-3: Deviant learner grammatical collocations

One thing which is instantly noticeable in Table 4-3 is that most errors in the learners' grammatical collocations involve the use (or, rather, misuse) of verbs and prepositions. This could be due to two reasons: First, the number of verb-centred grammatical collocations in English is far greater than those of noun or adjective-centred grammatical collocations. Second, Chinese is a verb-oriented language and students cannot seem to get rid of the habit of considering verbs first when writing English phrases. Unfortunately the students have not internalised enough grammatical collocation 'frames' (i.e. verb + NP + V-ing etc.) in English and either use a

non-existent frame or adopt the wrong frame for a particular verb.

For example, the students are inventing a grammatical collocation frame when they write *make our society can* and *make them can't* (*verb + NP + Aux). The original sentences containing this deviant structure follow:

(9)

a. I hope that government can think out one practicable way to stop this and pornography culture to **make our society can** [be] purified.
b. But many people still think [it] **make them can't** pay attention to drive and harmful to public morals.

Besides inventing grammatical collocations, the learners also impose 'wrong frames' of grammatical collocation on verbs which do not belong to the assigned category. For example, some of the CJU students write *avoid children touch* (verb + NP + verb-root), and *avoid their children looking* (verb + NP + verb-ing). In both cases, the grammatical collocation frame is not suitable for the verb *avoid.*

As we can see, the concept of grammatical collocation can cover plenty of areas between stand-alone lexical items and creative grammar. And it is worth noting that, in fact, there are only a limited number of grammatical collocation frames in English, and it will serve the students well, in a computer assisted writing/translating system, to provide the student writer with the corresponding frame when a particular verb is invoked.[24]

4.2 The translation corpus

A translation learners' corpus is different from a parallel corpus used in bilingual lexicography and machine translation research, as described in Kenny (1998). In these cases, the source texts and the corresponding target texts in the corpus act as good examples of translation from which norms in lexis and syntax are extracted to guide the formulation of bilingual lexicons or machine translation rules, etc. When the translation is done into a target language in which the translator is a learner, however, the corpus can no longer be used as norms for extraction. The corpus is instead a kind of error corpus (as it is named by Dodds 1999) from which deviant forms in the target language and bad examples of translation can be isolated and studied. The learner translation corpus also differs from the usual parallel corpus in another respect: For a common parallel corpus there is usually a one-to-one

[24] The automatic extraction of grammatical collocations should be easier to do than that of lexical collocations, since the content word and its function collocate are often adjacent, like *fight against*. In lexical collocations, for example, *take bribes*, words are likely to intervene, e.g. *take massive bribes*, which adds to the computational complexity.

relationship between the source text and the target text. For a learner translation corpus, on the other hand, the relationship is usually one-to-many, for the simple fact that the same source text is usually given by the instructor for many students to render translations based on it. Hence a learner translation corpus has the merit of offering different translation versions for the same source text, which is good for discovering different translation capabilities, styles and strategies, etc.

In the Department of Translation, CJU, translation texts are constantly being collected and added to the translation corpus. To illustrate the qualities and reveal the problems of the translations done by the CJU students, a sample translation corpus is again selected which consists of 42 English translations of the same Chinese text rendered by 42 subjects and amounting to approximately seven thousand words.

4.2.1 The basic and the idiomatic

Dodds (1999) correctly points out the primary difficulty of translating into a second language: namely, the translator's lack of proficiency in L2. Because of this limitation, Dodds advocates that emphasis should be laid on training the translation learners on the 'basics' of the target language, before encouraging them to use more creative skills. "Imaginative translations such as 'feelgood factor' and 'psychological climate' lose their positive impact when they appear in a context where basic errors are made" (p. 63), says Dodds when commenting on a translation excerpt. Dodds' words depict a sombre reality faced by all instructors teaching translation into L2, when he says that "the student has merely created a distortion of the English language in his failure to render tense, voice, lexis and even definite articles and capital letters according to the English norm" (ibid). To illustrate the fact that this phenomenon of 'defective basics' is also common to the CJU students' translations, an example is supplied below:

(10)

Teletubbies are the dolls that specially designed for the children before school age. They looks like people of the outer space very much, and there are different style of aerials were installed on their head. In addition, all of them have a little TV that helps them to get into contact with the external world. Through TV, teletubbies could have action together with children and could participate in children's games and activities.

In the magical-science island they have various movements that correctly reflect the truest behavior of children, and you will find that their behavior is mostly the same as your babies' whatever is singing, dancing and learning to speak.

The analysis of some of the errors in the translation excerpt follows:

Error	**Correction**	**Category**
the dolls	puppets	grammar & lexis
that specially designed	that are specially designed	grammar
They looks	They look	grammar
there are different style of aerials were installed	there are different styles of aerials installed	grammar
?get into contact with	get/keep in touch with	phraseology
Through TV	Through the television	grammar
?have action together with	interact with	phraseology
?have various movements	make various movements *or* perform various actions	phraseology
whatever is	whether it is	grammar & lexis
singing, dancing and learning to speak	singing, dancing or learning to speak	grammar

Table 4-4: Error analysis of the translation text in (10)

As can be seen from Table 4-4, very basic errors appear in the student's translation, like adding *-s* to a verb following a plural pronoun, omitting *be* or *the* where they should be present and adding them where they should not be, and using the inclusive conjunction (*and*) for the alternative one (*or*), etc.

On top of these basic grammatical errors, there are also unusual phrases like *have action together with* and *have various movements*, which may come from either a direct translation of the Chinese source, or a lack of knowledge in the phraseology of English. In fact, to improve students' grammar seems easier than to have them internalise plenty of prefabricated items like collocations and sentence builders. Examples of students' writings which are grammatically correct but unnatural in terms of selection of phrases or sentence patterns, are abundant. One example of this kind of writing in the sample translation corpus is:

(11)
The idols of Teletubbies are designed for children under the school age. They are like aliens. There are different aerials on their heads and there are televisions on their bellies. The Teletubbies can contact outside by the televisions. Through the televisions, Teletubbies can interact with other children and join in the games and activities of children.
Everything of Teletubbies on the teletubbyland can reflect the real actions of children. No matter singing, dancing or learning languages, you will find that they are very like your babies.

The passage in (11) reads very smoothly, i.e. there is little grammatical flaw to obstruct the reading of it. However, one can also feel the inadequacy in the text which comes from both the bad choice of vocabulary (*idols* rather than *puppets*) and phrases (*reflect the real actions* rather than *reflect the real behaviours*) and the monotonous sentence structures. One talks about 'inadequacies' instead of 'errors' in the case of learners' writings like (11). Peculiarly, it is in this area of inadequacies in terms of word/phrase/sentence structure selection where the learner needs the most help, probably more than they do in the area of errors in grammar. The analysis in the next section looks more closely at CJU students' ability in writing correct English sentences and in translating idiomatically into English, and the relationship between these abilities, if any.

4.2.2 The analysis

The sample translation corpus will be analysed in this section in the following fashion. Each of the 42 subjects' translations will be scrutinised in two respects: the basic features and the idiomatic features (to be explained later). Each subject will then have a set of two scores associated with her/his writing, one representing the basic abilities in grammar and other mechanicals like punctuation and spelling (hereafter the Basic score), the other representing the ability to write idiomatic English (hereafter the Idiomatic score). Whether these two sets of scores are correlated or

not will then be calculated. The hypothesis is that there is no correlation between these two scores. That is, better ability in English grammar does not necessarily guarantee equal advantages in idiomatic usage of English.

The Basic score for each subject is easy to determine. As if marking a student's examination papers, we will calculate the grammatical errors in each student's writing and simply deduct percentage five points from the 100% total score for each mistake made. The types of errors in grammar and their counts in the entire sample corpus are shown in Table 4-5. The statistics of the 42 subjects Basic scores are given in Table 4-6 and Figure 4-1.

Error type	**Example**	**Count**	**Percentage**
Omission/Addition	you will find they [**are**] just like your own baby (omission) that **is** reflects the kid's real behavior (addition)	91	41.0
Wrong word/form	They live **in** the Magical Technology Island (wrong preposition) Whether they are singing and dancing, or **learn** to speak (wrong verb form)	86	38.7
Word order	small size of televisions	16	7.2
Comparison	the most real behaviors	12	5.4
Punctuation	These things, which they do in the magically technological island just, reflect the true actions of children.	17	7.7
Total		222	100

Table 4-5: Types of grammatical errors and counts

Basic score (marks)	**Frequency (subjects)**
15	1
35	1

60	4
65	6
70	9
75	4
80	5
85	6
90	2
95	3
100	1
Total	42

Table 4-6: Frequency table of the Basic scores

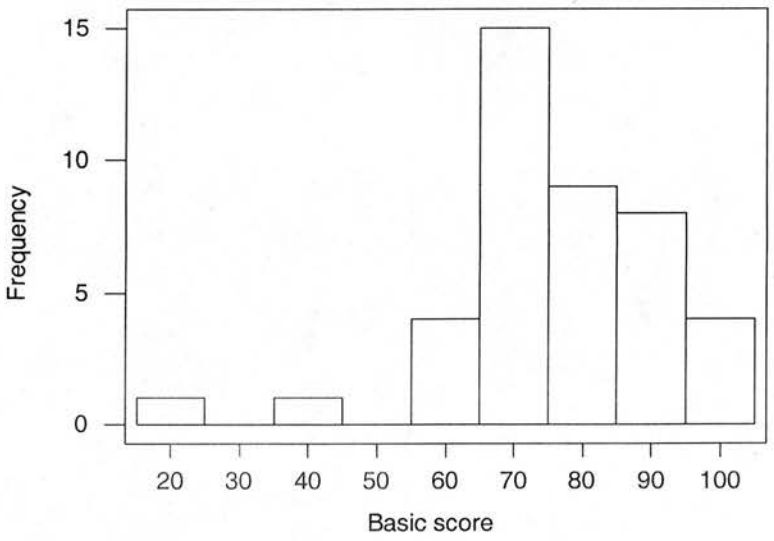


Figure 4-1: Histogram for Basic score

The Idiomatic score, on the other hand, is more difficult to decide, since it is not so easy to judge what is idiomatic and what is not. For the purpose of this study, a native speaker of English has been asked to supply a rewriting of the translation text to act as the norm against which we will judge the idiomaticity of the learner's writing. First we pick out several places in the native speaker's writing which best illustrate the idiomatic usage of the language, as the boldfaced parts in (12) show:

(12)
Teletubbies are puppets [**designed specifically**]$_1$ for [**preschool children**]$_2$. They [**look like**]$_3$ aliens with antennas on their heads. Each of them has a small television on their tummies through which they [**communicate**

with]$_4$ [**the rest of the world**]$_5$. [**Thanks to**]$_6$ the television, the teletubbies can [**interact with**]$_7$ children by [**participating in**]$_8$ their games and activities. Everything they do on the magic technology island reflects exactly the most genuine behaviour of the children. [**No matter whether**]$_9$ they are singing, dancing or learning to speak, you will find they [**bear remarkable resemblance**]$_{10}$ to your children.

Table 4-7 shows the category of each kind of word combination in (12):

Word combination	**Category**
designed specifically	lexical collocation (verb-adverb)
preschool children	lexical collocation (adjective-noun)
look like	grammatical collocation (verb-preposition)
communicate with	grammatical collocation (verb-preposition)
the rest of the world	lexical collocation (quantifier-noun)
thanks to	grammatical collocation (noun-preposition)
interact with	grammatical collocation (verb-preposition)
participate in	grammatical collocation (verb-preposition)
no matter whether	sentence builder
bear remarkable resemblance	lexical collocation (verb-noun & adjective-noun)

Table 4-7: Word combination types in (12)

The next step is to pick out the corresponding places in the students' writings and evaluate their idiomatic quality as compared with the native speaker's usage. For each of the ten items in Table 4-7, the highest mark is 10 (for a complete match), with the score decreasing as the resemblance diminishes. So for example for the native speaker's collocation *designed specifically*, the 42 subjects' corresponding use of words or phrases and the marks given are summarised in Table 4-8.[25]

Students' variant	**Number of subjects**	**Score**
designed especially	3	8
designed specially	2	8

[25] The scoring could have been done more objectively by cross checking with a native speaker, given more time in preparing this part of the thesis. The author, however, plans to write a separate article using this methodology to explore the same theme and to involve more native speakers in supplying text and making judgements.

especially designed	4	7
specially designed	9	7
particularly designed	2	6
only designed	1	5
designed[26]	19	3
especially set	1	2
produce	1	1
Total	42	

Table 4-8: Example marking of Idiomatic score for the item *designed specifically*

Note that the assigning of scores is reasonably objective although it is not completely mechanical, thanks to the clear native speaker model and the hierarchical structure of students' variants derivable from the model, from which the scores can also be hierarchically assigned. Thus in this way the ten items of native speaker's writing in Table 4-7 give rise to ten sets of Idiomatic scores to the 42 subjects like the set given in Table 4-8.[27] The statistics for the 42 subjects' Idiomatic scores are shown in Table 4-9 and the histogram follows as Figure 4-2.

Idiomatic score (marks)	**Frequency (subjects)**
44	2
45	1
46	1
47	2
48	1
51	1
53	2
54	3
55	1
56	4
57	2
58	1
59	1
60	2
61	2

[26] A surprising number of students undertranslate here, neglecting the meaning of 'specifically' in the source text altogether. This could be partially due to their lack of knowledge in the adverb-verb or verb-adverb collocations in English.

[27] See Appendix A for a full listing of the ten items from Table 4-7 used for the marking of Idiomatic scores.

62	3
63	2
64	3
65	3
68	1
69	1
72	1
74	1
80	1
Total	42

Table 4-9: Frequency table of the Idiomatic scores

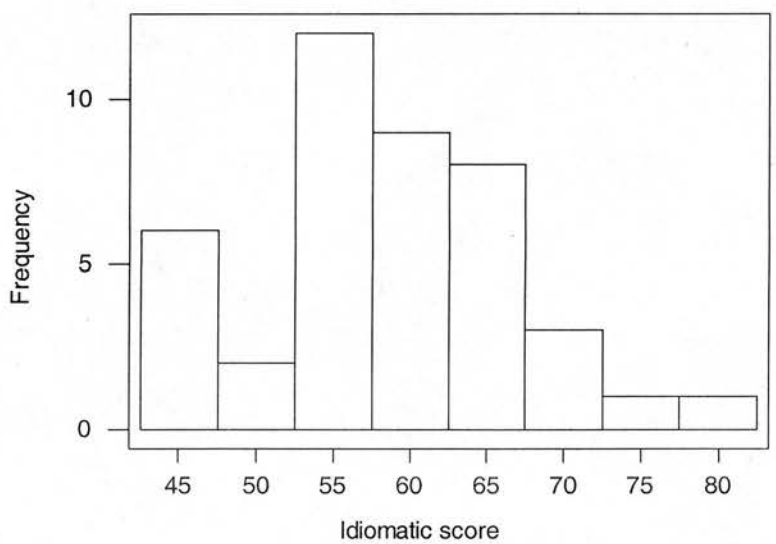


Figure 4-2: Histogram for Idiomatic score

Now we have a set of two scores for each subject which we can correlate. Table 4-10 shows each of the 42 subjects' scores and Figure 4-3 plots the correlation between the two variables.

Subject	B. score	I. score	Subject	B. score	I. score	Subject	B. score	I. score
1	95	72	**15**	35	56	**29**	80	74
2	80	47	**16**	15	47	**30**	95	62
3	95	63	**17**	80	58	**31**	70	56
4	75	55	**18**	70	54	**32**	60	46
5	70	60	**19**	70	45	**33**	85	62

6	65	65	**20**	80	60	**34**	70	48
7	75	53	**21**	65	69	**35**	70	57
8	75	65	**22**	60	57	**36**	65	62
9	65	56	**23**	90	59	**37**	65	68
10	90	56	**24**	80	64	**38**	85	54
11	85	64	**25**	70	44	**39**	60	61
12	85	80	**26**	70	64	**40**	65	54
13	70	53	**27**	75	63	**41**	100	61
14	85	51	**28**	85	65	**42**	60	44

Table 4-10: A set of two scores for each subject

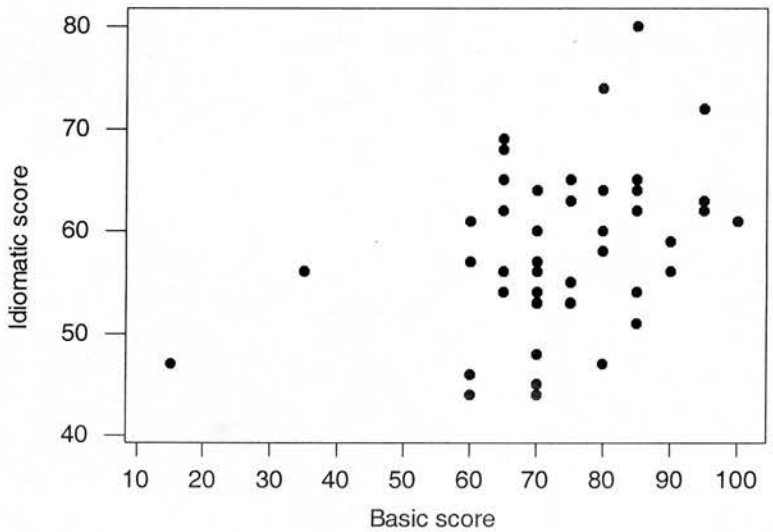


Figure 4-3: Correlation between Basic score and Idiomatic score

At a glance Figure 4-3 seems to show the relationship between Basic and Idiomatic scores to be a scatter plot. To calculate the correlation coefficient between the Basic scores and the Idiomatic scores, we use Spearman's ranked correlation test, and the result is:[28]

$$R = 0.329,\ N = 42,\ p <= 0.03514\ (Z = 2.1068)$$

[28] This is obtained from IFA Services at: http://www.fon.hum.uva.nl/Service/Statistics/RankCorrelation_coefficient.html, where the data in Table 4-10 are keyed in and the results are obtained automatically. Another similar set of results are obtained from VassarStats at http://faculty.vassar.edu/lowry/corr_rank.html, where r=0.3217, n=42, p<=0.037656, t=2.15.

The correlation coefficient $R = 0.329$ indicates that there is very low correlation between the Basic score and the Idiomatic score. In other words, in the population being considered, student ability in English grammar does not seem to be strongly correlated with student ability in writing idiomatic English.

4.3 Summary and conclusion

The discussion in this chapter centres around the theme that the analysis of collocation and other semi-fixed items in English is pedagogically compulsory and computationally tractable. Although the small scale analysis presented in this chapter seems to show that students' abilities in using grammar and to write idiomatic chunks are only weakly correlated, firmer results may only be obtained by larger scale experiments. But even if it turned out that these two abilities are strongly correlated, it should not be too surprising, since competences in aspects of language should supposedly move ahead together, even at different strides. However, we do not rule out the possibility that Chinese students' ability to use English collocations may still be particularly weak, falling far behind that of the native speaker.[29] According to Helen Pain (personal communication), in composition exercises "the most able students (who may learn more in the long term) are often more adventurous and experiment more with idiomatic use, sometimes at the expense of grades" and, on the other hand, the ordinary learner may "focus on being grammatically correct and avoid most idiomatic use". Since most Chinese students are of the conservative type in language classes, it may not be surprising that they will concentrate on being grammatically correct rather than on exploring the idiomatic domain.

In this chapter I also suggested that automatic analysis and correction of grammar in learners' writing is still not a viable option, but the analysis of less creative items, such as grammatical collocations, can be efficient and inspiring. This can not only help us to understand learners' competence in the second language and in translating into L2, it can also help us achieve the pedagogical goals of fostering better writers in the second language.

Another point worth noting from the discussions in this chapter is the use of a learner translation corpus to uncover the weaknesses in learners' writing of L2, which may be more revealing than a learners' composition corpus. For one thing, in translation the learner cannot avoid the use of certain vocabulary or structures which they have no confidence in (and which they would have chosen not to use in making a composition), as the learner's writing is restricted to the realm of the source text.

---

[29] Actually, a questionnaire survey in this connection is done and will be reported on in the next chapter.

Also, a group of students doing translation based on the same source text is bound to yield some interesting results, revealing different capacities and behaviour of the learners in using the second language, allowing the researcher to make comparisons and draw conclusions, as the discussion in this chapter hopefully shows.

In Chapter Three a model of translation competence was proposed which asserts that such competence must be built on competence in the second language. In order to understand what the Chinese learners' competence is like in English as a foreign language, in this chapter we set out to examine some Chinese students' English compositions and English translations of Chinese texts. We noted down some of the Chinese learners' inadequacies in using English lexical or grammatical collocations. In the next chapter, we will more systematically compare Chinese learners' competence in English collocations with that of native speakers. It is hoped that these findings will provide motives for exploring deeper into the teaching of collocation and other idiomatic aspects in English. While suggesting that the computer is a good environment for integrating translation and second language teaching, this thesis also emphasises the acquisition of idiomatic competence in such a favourable environment, arguing for the inseparability of idiomaticity and good translation.

Chapter 5

Collocations and Lexical Phrases

In the last chapter we investigated CJU students' competence in English grammar and idiomatic usage as well as in the middle ground of grammatical collocations. In this chapter we will look deeper into the phraseology of English, investigating the nature of collocations and the so-called lexical phrases and their relevance to language processing, language acquisition, etc. In particular, we will do a systematic study of the differences between native speakers' and the L2 learners' abilities to use English collocations. Finally, we will see how these findings tie in with the model of translation competence we developed earlier, and propose ways in which to achieve the pedagogical goals of strengthening learners' capacities to use English collocations and other phrasal combinations.

5.1 Rule-based vs. Instance-based Processing

Sinclair (1991:109-110) makes a distinction between the open-choice principle and the idiom principle in interpreting texts from corpora. The open-choice principle applies when the writer/speaker creates utterances from a complex set of possibilities which is constrained only by grammar. The idiom principle applies when the writer/speaker chooses from a large number of pre-constructed phrases. Like the repetitive nature of daily human affairs, Sinclair suggests that the idiom principle is the first to be applied in the use of language, and the user switches to the mode of open-choice only "when there is a good reason" (p. 114). The distinction Sinclair made has since been widely quoted in the linguistic literature and his view on cognitive processing has also been supported by some research findings. Skehan (1998), for example, explains the example based theory (corresponding to Sinclair's idiom principle) on language processing as an alternative to the previously dominating theories of rule-based systems (corresponding to the open-choice principle). In this connection, Skehan brings to our attention the important notion of 'native-like selection' originally proposed by Pawley & Syder (1983), also widely quoted, who propose that even though non-native speakers can speak or write L2 fluently, they are still perceived by native speakers to be "operating a different system" because of their non-standard choice of language. Referring back to Sinclair's definitions, we can say that non-native speakers have acquired a good command of grammar in L2 to satisfy the needs of open-choice processing mode, but

they are seriously lacking in the collection of prefabricated phrases which accounts for the failure in idiomatic processing mode.

Pawley and Syder, according to Skehan, propose that one of the things that constitute native-like fluency is the lexicalised sentence stems (LSS). For example, *NP be-TENSE sorry to keep-TENSE you waiting* (e.g. *I'm sorry to have kept you waiting*) is an LSS, and an average native speaker knows "hundreds of thousands" of LSS' like this. A more comprehensive treatment of the subject can be found in Nattinger & DeCarrico (1992), who discuss 'lexical phrases' of all kinds, among which 'sentence builders', like *there's no doubt that X* or *my point here is X*, are similar to Pawley and Syder's lexicalised sentence stems. Nattinger & DeCarrico talk about 'pragmatic competence' as opposed to grammatical competence, which they say is the ability to select the correct pre-assembled forms to use in a certain situation. According to Nattinger & DeCarrico, prefabricated items allow the language user to speak fluently, as they make it easier to retrieve larger chunks of lexical units, rather than having to monitor the production of individual words, which allows the speaker to pay more attention to discourse structure. Nattinger & DeCarrico quote Becker's (1975:72) idea on the processing of speech to strengthen their point:

> We start with the information we wish to express or evoke, and we haul out of our phrasal lexicon some patterns that can provide the major elements of this expression. Then the problem is to stitch these phrases together into something roughly grammatical ... and if all else fails to generate phrases from scratch ...

Beck's view above, incidentally, is exactly the same as Sinclair's view discussed earlier that the idiom principle plays the dominant role in language processing, with the open-choice option being evoked only when there is a good reason. Thus, it should be safe to conclude this discussion by saying that prefabricated language strings of all kinds probably play a much heavier role in language processing and acquisition than people like to think. Although, in principle, language is creative and grammar is 'generative', yet, as Widdowson (1989) amply points out, "not all access is dependent on analysis". The implication for language teaching is that, although the instruction of grammatical rules is indispensable, yet perhaps we should take hints from Sinclair to decide the place of grammar in pedagogy, who says, "The open-choice analysis could be imagined as an analytical process which goes on in principle all the time, but whose results are only intermittently called for". My interpretation of these words, in terms of language teaching, is that grammar is basic, but pre-assembled chunks are even more important for language learners. This happens to be Widdowson's (1989) conclusion to his own famous article, who says, "

Linguistics may be about grammar rather than language. But the study and teaching of language is about a lot more than that, and grammar needs to be put in its place." (p.136)

5.2 Conventionalised forms

In this section all the categories of habitual word combinations, other than collocations, are considered together, allowing the latter to be discussed separately. Because Nattinger & DeCarrico (1992) explicitly distinguish collocations from lexical phrases, it is convenient to follow their approach and discuss all conventionalised forms other than collocations under the rubric of Nattinger & DeCarrico's lexical phrases, which are "form/function composites, lexico-grammatical units that occupy a position somewhere between the traditional poles of lexicon and syntax..." (p. 36). Nattinger & DeCarrico recognise four kinds of lexical phrases:

- Polywords: Short phrases which function like individual words. Constituent parts are not variable. e.g. *by the way, so long, as it were.*
- Institutionalized expressions: Sentence-long fixed units which usually function as a separate utterance. e.g. *nice to meet you, give me a break.*
- Phrasal constraints: Short to medium phrases allowing internal variations. e.g. *see you* ____ *(soon, later ...), as far as I* ____ *(know, can tell ...).*
- Sentence builders: Framework for constructing whole sentences. e.g. *let me start by* ____ *(mentioning a few important things, pointing out a truth ...).*

What is important to this thesis is Nattinger & DeCarrico's interesting distinction between collocations which are lexical phrases and collocations which are not lexical phrases. According to them:

> Prefabricated phrases are collocations if they are chunked sets of lexical items with no particular pragmatic functions; they are lexical phrases if they have such pragmatic functions. (p. 37)

Thus for example in written discourse, based on Nattinger & DeCarrico's scheme, *in most cases* and *to sum up* are lexical phrases, since they have the 'pragmatic functions' of 'generalising' and 'closing up' respectively; whereas collocations like *commit suicide* and *readily available* are not lexical phrases since they are not assigned any pragmatic functions. This is of course debatable, since we cannot say that *readily available* used in a sentence like *Parking spaces should be made readily available for the disabled* has no pragmatic function at all, as the sentence hosting it

must have some pragmatic function (like 'asserting') in the discourse and the collocation as part of the sentence is bound to have some sort of contribution to this function. It is probably safer to say, instead, that some collocations have independent pragmatic functions (e.g. *to sum up, in general*) and some are the main carrier of pragmatic functions (e.g. *it can be claimed that..., I strongly disagree with...*); while others only give secondary support of various degrees to the pragmatic function of the sentence they reside in (e.g. *readily available, cause distress*).

If we make the above distinction, then it becomes clear that what Nattinger & DeCarrico are concerned about are only the first two types, i.e. those phrases or sentence stems which have independent or primary pragmatic functions in discourse. However, these cannot be the only kinds of collocations or fixed expressions that the L2 learners need to know. Nattinger & DeCarrico's suggestion of teaching students to use structures like the following is certainly sound and useful:

> *For a long time, many researchers have believed X. This paper will show that Y, by comparing and contrasting Z, and recommending that A.* (p. 170)

But between these elements of skeletons, there is lots of flesh to be filled in, and this will contain some collocations (e.g. *growing economy, sorely needed*) and some sentence stems (e.g. *it is difficult for...to..., no...could be more...*) which have only weak pragmatic functions in terms of discourse structures and are not subsumed in Nattinger & DeCarrico's discussions. These will have to be taught to the learners as well in order to complete the whole task of writing.

It is my observation, as an EFL practitioner for some years, that some of the lexical phrases in Nattinger & DeCarrico's sense and the general theory behind them (i.e. the pragmatic function, or using them as signifiers of discourse structures in the context of writing) have already caught the attention of practitioners and textbook writers in the field, as manifested in the textbooks of the rhetoric approach (e.g. Ruetten 1997: 84-85), the 'paragraph development' approach (e.g. Arnaudet & Barrett, 1981) or the 'connecting and combining' approach (e.g. Mills, 1982) to English composition. What the field of ESL is still lacking in at this stage is actually the introduction of the kind of collocation excluded by Nattinger & DeCarrico -- the collocations that are not assigned a pragmatic function, or in the terms used here, that do not have stand-alone structure-signalling functions in written discourse. The following section is concerned with these collocations, since these have received relatively little attention in the field of second language teaching.

5.3 Collocation

Second language learners frequently make grammatically sound sentences which are nevertheless unacceptable or unnatural. One reason is that they have not internalised enough knowledge about collocation in English. Collocations are even more difficult to master than idioms. Idioms are mostly closed sets, and learners can simply memorise one as a lexical unit, but not so with collocations. Collocation is a productive process, albeit a limited productive process. What is more confusing is that in some cases there are sometimes observable rules governing which words can collocate with which other words. For example, Altenberg 1991 finds that 'maximizers' like *completely* can modify non-gradable adjectives like *wrong*, but not gradable adjectives like *big*. However, more often than not, there is no easily inferable rule (for example, why *achieve a level*, but not *achieve a point*?). This section discusses the nature of English collocations and investigates second language learners' knowledge of some types of English collocations and the possible difficulties they face in mastering collocation.

5.3.1 The nature of collocation

Stubbs (1995) defines collocation as 'the habitual co-occurrence of words'. Though here, remember, we are focussing specifically on non-pragmatic (or at most secondary) use. But how does a collocation like *cause the problem* differ from an idiom like *kick the bucket*, both appearing to comply with the definition?

Cruse (1986) usefully points out that for a collocation, each lexical constituent is also a semantic constituent, so in *cause the problem*, *cause* and *problem* retain their original meaning. In the case of idioms, however, all lexical constituents form a single semantic unit, so *kick* and *the bucket* combine to form the meaning of 'die', each losing its individual meaning. However, as Cruse readily notices, there is nevertheless some sort of cohesion existing among the constituents of a collocation; otherwise there would be no difference between collocations and free combinations. So for example, Cruse observes that it is the notion of 'consumption' which constrains the collocation of *heavy* with *drinker, smoker, drug-user* etc. Stubbs further explains the mechanism of 'semantic prosody' underlying the formation of some collocations.

For example, the semantic prosody for *cause* collocating with nouns will be one pointing to unpleasant things, like *cause problems* and *cause unhappiness*, etc. The semantic prosody for *provide*, on the other hand, will be a positive one, like *provide food*. But semantic prosody certainly can not explain all types of collocations, for example, why *extremely helpful* but not *extremely right* (a Chinese ESL learner's phrase), both having the same positive semantic prosody? Different types of collocations seem to require different explaining strategy, and more work is definitely

needed to be done on exhaustive classification of English collocations and on the various combining principles underlying them.

5.3.2 An Investigation

In order to understand ESL learners' ability in using English collocations, the author has conducted a questionnaire survey. The questionnaire contains fifteen questions, each of which presents four candidate verbs followed by a noun in the context of a sentence, as follows:

> He said he had already _____ a complaint.[30]
> ( )lodged ( )made ( )reported ( )submitted

All the sentences (i.e. questions) in the questionnaire have been carefully adapted from the hundred-million-word British National Corpus (BNC). Context (i.e. the sentence) is kept as minimal as possible so that, it is hoped, it does not favour the choice of a particular word. All four candidate verbs in each question are carefully chosen too. The main criterion is that they be synonyms to each other. WordNet (n.d.) is consulted first to this end. Where it is not possible to do so, words are chosen which do not essentially change the proposition of the sentence. In the case that there are more than four eligible candidates, those are chosen which maintain as larger a gap in z-score (more on this later) between each other as possible, so that, it is hoped, their different degrees of collocatability will more likely be reflected by the human rating.

In the experiment the subject is asked to rank the suitability of the verbs in each question. For each question there is a 'standard answer' calculated from the BNC with the z-score method.[31] So for the above example, calculations from BNC yield the z-scores of 113.78, 21.86, 0.45, and 5.5 for *lodged complaint, made complaint, reported complaint*, and *submitted complaint* respectively. The best collocation in this set is thus *lodged complaint*, the second best *made complaint* and so on, the higher the z-score, the stronger the collocational bond, and the subject's performance is evaluated according to this corpus-extracted norm.

There were three groups of subjects involved in this study. The control group consisted of 119 native speakers of English. There were two other groups: a Chinese learners group (42 subjects) and a European language speakers group (32 subjects).

---

[30] See Appendix B for the questionnaire in its entirety.

[31] See Barnbrook (1996) for a detailed description of how to calculate z-scores for word collocations from corpus.

All the participants were current UK university undergraduates or postgraduates at the time of surveying, and the questionnaires were delivered and answered in the form of e-mail. Although the questionnaire did not specifically test or ask about the subjects' general proficiency in English, it made the reasonable assumption that all the university students answering the questionnaire who considered themselves native speakers are adult speakers who used English as their first language. As for students who considered themselves non-native speakers of English, since they were studying in a UK university, which generally requires an entering international student to pass a certain level for a recognised English proficiency test (e.g. TOEFL 550), it was assumed further that all the non-native participants were at least higher intermediate to advanced learners of English as a second or foreign language.

The results discussed below involve firstly an informal comparison between the 'standard answers' from BNC and the performance of native speakers of English, and then tests for significance in differences among the performances of the three groups of participants.

Noun and Candidate Verbs	**z-scores from BNC and ranking**	**Native speakers collective ranking**
1. complaint ( )lodged ( )made ( )reported ( )submitted	1(113.7) 2(21.8) 4(0.4) 3(5.5)	1(37%) 2(36%) 4(7%) 3(20%)
2. activity ( )do ( )execute ( )perform ( )practise	4(-9) 2(-0.4) 1(3.8) 3(-0.6)	2(26%) 4(12%) 1(44%) 3(18%)
3. issue ( )address ( )cover ( )examine ( )treat	1(75) 4(-0.4) 2(6.2) 3(0.2)	1(39%) 3(18%) 2(33%) 4(9%)
4. confidence ( )acquire ( )gain ( )get ( )obtain	2(4.1) 1(63.6) 4(-0.6) 3(1.5)	2(29%) 1(45%) 4(12%) 3(13%)
5. information ( )convey ( )deliver ( )transfer ( )transmit	1(53.3) 3(9.1) 4(-2.6) 2(29.8)	1(44%) 2(27%) 3(15%) 4(14%)
6. relationship ( )build ( )construct ( )establish ( )make	2(11.7) 3(0.6) 1(38.2) 4(-1.2)	2(36%) 3(14%) 1(43%) 4(7%)
7. response ( )arouse ( )elicit ( )extract ( )stimulate	3(2.9) 1(92.7) 4(0.9) 2(8.6)	3(18%) 1(41%) 4(17%) 2(25%)
8. trouble ( )cause ( )generate ( )make ( )stir-up	2(64.4) 4(-0.7) 3(6.4) 1(128.1)	1(38%) 4(6%) 3(27%) 2(29%)
9. agreement ( )achieve ( )get ( )make ( )reach	2(10) 4(1.3) 3(1.7) 1(124.2)	2(27%) 3(11%) 4(8%) 1(54%)
10. research ( )do ( )conduct ( )perform ( )undertake	4(-12) 2(49.6) 3(2.8) 1(58.4)	3(17%) 1(39%) 4(16%) 2(28%)
11. experience ( )acquire ( )get	2(7) 4(-2.5) 1(63.3) 3(1.9)	2(25%) 4(16%) 1(40%) 3(19%)

( )gain ( )obtain		
12. example ( )cite ( )make ( )quote ( )take	1(60.5) 4(-1.7) 3(7.8) 2(35.8)	1(40%) 4(11%) 2(30%) 3(19%)
13. service ( )conduct ( )execute ( )perform ( )provide	3(6.7) 4(-0.6) 2(16.7) 1(74.8)	3(11%) 4(6%) 2(31%) 1(53%)
14. knowledge ( )absorb ( )acquire ( )learn ( )obtain	3(-0.6) 1(74.5) 4(-0.8) 2(5.8)	4(7%) 1(46%) 3(8%) 2(38%)
15. success ( )assure ( )ensure ( )guarantee ( )warrant	3(2.6) 1(32.8) 2(31.8) 4(-0.6)	3(11%) 1(42%) 2(39%) 4(7%)

Table 5-1: Comparison between corpus statistics and native speakers' intuition

Table 5-1 compares native speakers' performance with the 'standard answers' calculated from BNC. In the BNC column, the z-score for each verb appears in its corresponding position in relation to other verbs in the same set, in parenthesis directly following its ranking in the set. The 'z-score for each verb' should be understood to mean the z-score for that verb collocating with the noun supplied in that particular question.

The subjects' ranking column needs more explanation. In the questionnaire the subjects were asked to rank the verbs from 1 to 4 (see Appendix B for details). In making the above table, the ranking was converted to scores (1 gets 4 marks, 2 gets 3, 3 gets 2, and 4 gets 1), adding each verb candidate's scores up for the whole group, and ranking them again to get the above result.

So in the Native speakers column in Table 5-1, the ranking is determined by the collective scores given by the 119 subjects to each verb. The percentage of each verb reflects the scores it receives in relation to the set of four verbs in each question. So for Question One, for example, the percentages mean approximately that the subjects in this group collectively give the score of 37% to the verb *lodged* as to its suitability to appear with the noun *complaint* in that sentence (making it the best choice in this set for this group). Likewise, 36% is given to *made*, 7% to *reported* and 20% to *submitted*.

The two sets of ranking show remarkable similarity in Table 5-1. We see exactly the same ranking in seven out of fifteen questions (1, 4, 6, 7, 11, 13, 15) in the BNC and the native speakers' columns. In three other questions (3, 9, 14) the same first and second choices are maintained. For the remaining five questions, three (2, 5, 12) have the same first choice in the two columns. Even in the two most dissimilar questions (8, 10), the first two choices are correctly singled out from the lower ranked ones (third and last) disregarding the fact that their order is reversed -- Thus the corpus evidence and the native speakers intuition largely coincide.

Non-native speakers' scores in Table 5-2 and Table 5-3 below, however, tell another story.

Noun and Candidate Verbs	**z-scores from BNC and ranking**	**Chinese speakers collective ranking**
1. complaint ( )lodged ( )made ( )reported ( )submitted	1(113.7) 2(21.8) 4(0.4) 3(5.5)	4(9%) 1(48%) 2(25%) 3(18%)
2. activity ( )do ( )execute ( )perform ( )practise	4(-9) 2(-0.4) 1(3.8) 3(-0.6)	1(34%) 4(18%) 2(29%) 3(20%)
3. issue ( )address ( )cover ( )examine ( )treat	1(75) 4(-0.4) 2(6.2) 3(0.2)	1(37%) 2(30%) 3(28%) 4(5%)
4. confidence ( )acquire ( )gain ( )get ( )obtain	2(4.1) 1(63.6) 4(-0.6) 3(1.5)	4(15%) 1(35%) 2(29%) 3(21%)
5. information ( )convey ( )deliver ( )transfer ( )transmit	1(53.3) 3(9.1) 4(-2.6) 2(29.8)	2(30%) 1(39%) 4(13%) 3(19%)
6. relationship ( )build ( )construct ( )establish ( )make	2(11.7) 3(0.6) 1(38.2) 4(-1.2)	2(33%) 4(10%) 1(36%) 3(21%)
7. response ( )arouse ( )elicit ( )extract ( )stimulate	3(2.9) 1(92.7) 4(0.9) 2(8.6)	2(28%) 3(27%) 4(8%) 1(38%)
8. trouble ( )cause ( )generate ( )make ( )stir-up	2(64.4) 4(-0.7) 3(6.4) 1(128.1)	2(34%) 4(10%) 1(39%) 3(17%)
9. agreement ( )achieve ( )get ( )make ( )reach	2(10) 4(1.3) 3(1.7) 1(124.2)	2(25%) 4(16%) 3(22%) 1(38%)
10. research ( )do ( )conduct ( )perform ( )undertake	4(-12) 2(49.6) 3(2.8) 1(58.4)	2(32%) 1(34%) 4(6%) 3(29%)
11. experience ( )acquire ( )get ( )gain ( )obtain	2(7) 4(-2.5) 1(63.3) 3(1.9)	4(16%) 2(27%) 1(31%) 3(26%)
12. example ( )cite ( )make ( )quote ( )take	1(60.5) 4(-1.7) 3(7.8) 2(35.8)	2(27%) 1(28%) 3(26%) 4(20%)
13. service ( )conduct ( )execute	3(6.7) 4(-0.6) 2(16.7) 1(74.8)	3(15%) 4(11%) 2(17%) 1(57%)

( )perform ( )provide		
14. knowledge ( )absorb ( )acquire ( )learn ( )obtain	3(-0.6) 1(74.5) 4(-0.8) 2(5.8)	4(6%) 1(34%) 3(25%) 2(35%)
15. success ( )assure ( )ensure ( )guarantee ( )warrant	3(2.6) 1(32.8) 2(31.8) 4(-0.6)	3(16%) 2(37%) 1(38%) 4(9%)

Table 5-2: Comparison between corpus statistics and Chinese speakers' intuition

Noun and Candidate Verbs	**z-scores from BNC and ranking**	**European language speakers collective ranking**
1. complaint ( )lodged ( )made ( )reported ( )submitted	1(113.7) 2(21.8) 4(0.4) 3(5.5)	3(18%) 1(41%) 4(16%) 2(26%)
2. activity ( )do ( )execute ( )perform ( )practise	4(-9) 2(-0.4) 1(3.8) 3(-0.6)	2(30%) 3(18%) 1(44%) 4(9%)
3. issue ( )address ( )cover ( )examine ( )treat	1(75) 4(-0.4) 2(6.2) 3(0.2)	1(35%) 3(21%) 2(33%) 4(11%)
4. confidence ( )acquire ( )gain ( )get ( )obtain	2(4.1) 1(63.6) 4(-0.6) 3(1.5)	2(26%) 1(43%) 3(16%) 4(15%)
5. information ( )convey ( )deliver ( )transfer ( )transmit	1(53.3) 3(9.1) 4(-2.6) 2(29.8)	1(37%) 2(25%) 3(19%) 4(18%)
6. relationship ( )build ( )construct ( )establish ( )make	2(11.7) 3(0.6) 1(38.2) 4(-1.2)	2(35%) 3(10%) 1(49%) 4(6%)
7. response ( )arouse ( )elicit ( )extract ( )stimulate	3(2.9) 1(92.7) 4(0.9) 2(8.6)	4(8%) 2(39%) 3(12%) 1(41%)
8. trouble ( )cause ( )generate ( )make ( )stir-up	2(64.4) 4(-0.7) 3(6.4) 1(128.1)	1(43%) 4(7%) 2(34%) 3(16%)
9. agreement ( )achieve ( )get ( )make ( )reach	2(10) 4(1.3) 3(1.7) 1(124.2)	2(34%) 4(5%) 3(11%) 1(49%)
10. research ( )do ( )conduct ( )perform ( )undertake	4(-12) 2(49.6) 3(2.8) 1(58.4)	2(33%) 1(36%) 4(9%) 3(23%)
11. experience ( )acquire ( )get ( )gain ( )obtain	2(7) 4(-2.5) 1(63.3) 3(1.9)	3(23%) 2(28%) 1(35%) 4(15%)
12. example ( )cite ( )make ( )quote ( )take	1(60.5) 4(-1.7) 3(7.8) 2(35.8)	1(33%) 3(21%) 2(32%) 4(14%)
13. service ( )conduct ( )execute	3(6.7) 4(-0.6) 2(16.7) 1(74.8)	3(7%) 4(4%) 2(17%) 1(72%)

( )perform ( )provide		
14. knowledge ( )absorb ( )acquire ( )learn ( )obtain	3(-0.6) 1(74.5) 4(-0.8) 2(5.8)	3(4%) 1(51%) 4(3%) 2(42%)
15. success ( )assure ( )ensure ( )guarantee ( )warrant	3(2.6) 1(32.8) 2(31.8) 4(-0.6)	4(10%) 2(34%) 1(46%) 3(9%)

Table 5-3: Comparison between corpus statistics and European language speakers' intuition

Table 5-2 and Table 5-3 show collective ranking scores from 42 Chinese speakers and 32 other non-native speakers (mostly European language speakers) respectively in comparison to BNC. All are advanced learners studying in a UK university at the time of the survey. A careful examination of Table 5-2 shows a quite noticeable discrepancy between Chinese learners' collocational knowledge and the standard inferred from BNC. There are three questions (1,7,8) in which the verb ranked the best by BNC is marked by Chinese learners as the last or third choice. Conversely, there are four questions (2, 5, 8, 12) in which the lower ranked verb in BNC (number four or three) is ranked by Chinese learners to be the best. Not a single example of these occurs in the native speakers' ranking.

Finally, a casual inspection of Table 5-3 seems to show that the European language speakers' performance comes somewhere between that of the native speakers and that of Chinese speakers. Indeed, if we count the correctness of the first choice for the three groups in relation to the BNC we can see that the native speaker group has 13 correct first choice when compared with BNC, the European non-native speaker group has 10 correct choices, whereas the Chinese speaker group has only 7.

Another way to look at this is to sum the difference in rankings for each word, across all words. So, referring to Table 5-1, for the first word the difference in rankings between corpus statistics and native speakers is (1-1) +(2-2)+(4-4)+(3-3) = 0 (the rankings are all the same). For the second word, the difference is (4-2) +(4-2)+(1-1)+(3-3) = 4. The total difference in their rankings is 22. The difference in rankings however for European speakers and Chinese speakers are 42 and 56, respectively.

Overall, the above discussion suggests that, firstly, the methodology used here for eliciting collocation knowledge seems robust, as we use the same sets of test phrases to get very similar results from two different sources of standard English usage: the corpus evidence and the native speakers' intuition. Secondly, the collective ranking analysis seems to show that the Chinese learners' ability in judging English collocations is noticeably inferior to that of the native speakers', most likely

even worse than the other non-native speakers of English like the European language speakers.

While the above simple collective ranking analysis allows us to make a fast and informal comparison between the corpus standard and the performance of different groups, it also has the merit of letting us see clearly for which questions the groups differ in which way. We still need, however, more rigorous statistical tests to decide whether the three groups of subjects differ from each other significantly in their answers.

To conduct a more formal statistical test, first we need to convert the ranking to scores for each individual's reply to the questionnaire. The following marking scheme is used for this purpose:

Score	*Ranking made*					
BNC answers		*1*	*2*	*3*	*4*	*0*
	1	**5**	**4**	**1**	**0**	**-2**
	2	**4**	**5**	**2**	**1**	**-1**
	3	**1**	**2**	**3**	**2**	**1**
	4	**0**	**1**	**2**	**3**	**2**

So if the corpus says 'the answer for this verb should be 1' and indeed if the subject answers 1 for this verb, then he gets 5 marks. If he answers 2 instead, then he gets 4 marks. If he neglects this word (choosing 0), then he is penalised by being deducted 2 points, and so on and so forth. Following this scheme, the scores for the standard answers from BNC would be: (5+5+3+3)×15=240, which is the full mark one can get from this questionnaire.

Now if we calculate the scores for the three groups (as groups, not as individuals, which we would do later) in question based on this scheme, using the group ranking orders from Table 5-1, 5-2, and 5-3 respectively, we obtain:

Native speakers: 200
Non-native speakers (Chinese): 166
Non-native speakers (European): 184

The results should not be surprising. It stands to reason that native speaker score (200) is closest to the corpus norm (240), followed by the European learners group (184), with the Chinese learners group being the last (166). This only conforms to our earlier speculations in the informal discussion.

With the above marking scheme, we are able to compare the performance of individuals in the groups, in relation to each other. Since from the previous discussion we have evidence to suggest that there are differences between the groups, we can now support this claim by looking at the scores of individuals within these groups.[32] We can now do pairwise *t*-test to test the following null hypotheses:

1. There is no difference between the rankings of native and non-native (Chinese) speakers.
2. There is no difference between the rankings of native and non-native (European) speakers.
3. There is no difference between the rankings of non-native (European) and non-native (Chinese) speakers.

The results of Student t-test for Group A (native) and Group C (Chinese) are as follows:

#A = 119, #B = 42, t = 11.29, DF = 159, p <= 1.675e-21

Since t = 11.29 is past the critical value 1.645 at 0.95 percentile, we can reject the first null hypothesis and conclude that there is a difference between the rankings of native speakers and Chinese learners of English.

The results of the other two tests are as follows:

#A = 119, #B = 32, t = 4.857, DF = 149, p <= 3.053e-06
#B = 42, #B = 32, t = 3.845, DF = 72, p <= 0.0002635

Both t scores are past the critical values 1.645 (DF > 75) and 1.666 (DF = 72) respectively. Thus we can also reject the second and the third null hypotheses, after rejecting the first one. This amounts to saying that there are differences between all three groups' rankings. Thus the three groups of native speakers of English, non-native speakers of English (European), and non-native speakers of English (Chinese) each is operating a different system when doing the collocation test. From a comparison of the means of their ranking scores, we can further note the stronger ones and the weaker ones:

Group	N	Mean
Native	119	164.08

[32] Details of the individual scores for each group of subjects are provided in Appendix C

Chinese	42	122.24
European	32	143.88

As we can see from the comparison of means, native speakers are the most able in making collocation judgements, Chinese learners are the weakest, and European learners take the middle ground.

5.3.3 Discussion

Why there should be such differences between these three groups in making collocation judgements is an interesting question. All overseas students are requested by most UK universities to pass a certain threshold in their English proficiency (for example, TOEFL score 550) before they can come to study. The Chinese students participating in this study (who are currently studying in UK universities) should have reasonably good command of English to enable them to understand lectures, converse with the instructors, take exams, etc. in the university. Yet their abilities to use English collocations still fall considerably behind those of the native speakers. The European learners did better than the Chinese learners, but there is still a difference between their collocational abilities and those of the native speakers'. Disregarding their abilities in fluency or grammaticality of English, the two groups of learners are not on the same scale with native speakers at least in their collocational abilities.

There are many possible reasons to account for these differences. For example, this may be an inevitable consequence of communicative language teaching, which emphasises meaning and interaction and does not demand accuracy or refinement of written language. This may be due to learners' lack of motivation to improve their collocational abilities once they feel their communicative abilities are adequate for them to carry out intended functions. Or they may not be aware of such a thing as collocation at all – neither the teachers nor the textbooks have reminded them of such a phenomenon in English. Since the learners have not internalised enough language data to allow collocation to become automatic (as native speakers have), and since their attention has not especially been drawn to it, neither implicit nor explicit learning of collocation apply, hence their poor performance in collocation judgements. The verification of all these and other surmises requires further research which is beyond the scope of this thesis.

As for the difference between the European learners and the Chinese learners, there may be many possible reasons to account for it as well. For example, the European group possibly did better because of the similarity between their mother tongue and English, which facilitates positive transfer. They may have been

studying English longer than the Chinese learners do (a factor the questionnaire did not probe into). It may simply be due to their greater exposure to English in Europe, or because the European methodologies for teaching English are more effective for them.

The trace of mother tongue interference can be spotted in the Chinese learners' rating of English collocations. For example, in Chinese the verbs most frequently taking 'knowledge' as complement are the Chinese equivalents to *learn* and *absorb*, which are thus marked by Chinese speakers as preferred collocates for *knowledge* as well in English, not knowing that, in fact, *acquire* and *obtain* are the more natural collocates for *knowledge* in English, as evidenced by BNC and the native speakers' marking.

The 'generic verb' effect mentioned before is also possibly a factor here – When the Chinese learners are unsure which verbs to choose, primitive or generic verbs like *do, make*, etc. are good candidates to pick, probably simply because learners are more familiar with these verbs thanks to their higher frequency in the language in general than the other verbs.[33]

Although it seems that collocational errors do not necessarily cause misunderstanding, they do make learners' writing sound unnatural (like *make action, perform help*) or unacceptable (like *cease confidence*). What's more, the lack of the use of familiar English collocations (like *abundantly clear, readily available*) and the replacement of them with more clumsy paraphrases makes learners' writing sound foreign or immature. As Howarth (1998) advises, "Knowledge of arbitrary restriction on collocations is required in order to conform to the expectations of the academic community". The worst scenario for learners is when they "fail to communicate a clear meaning as a result of the unnaturalness of the lexical co-occurrence" (ibid.).

In view of the Chinese learners' weaknesses in English collocations, it is not hard to imagine that collocation will constitute a problem in translating into English for Chinese translators, no matter how fluent and grammatical their English texts are. One of the assertions of this thesis is for translation teaching and second language learning to converge in the same teaching environment. Likewise, translation educators can also collaborate with applied linguists to find solutions to the problems of collocation for translators as second language learners. Things which need to be done include a more exhaustive classification of collocations and the mechanisms governing their formation, a deeper understanding of the difficulties faced by learners

[33] In investigating learners' 'amplifier' collocations, Granger (1998) makes a similar observation that learners tend to use 'general-purpose' items like *very* to modify adjectives.

in mastering collocation, and a sound methodology for presenting collocation materials to learners, etc.

5.4 An instruction framework

Following the above investigation and discussions, we propose a model for teaching translation into a second language, particularly for Chinese students. This is embodied in Figure 5-1:

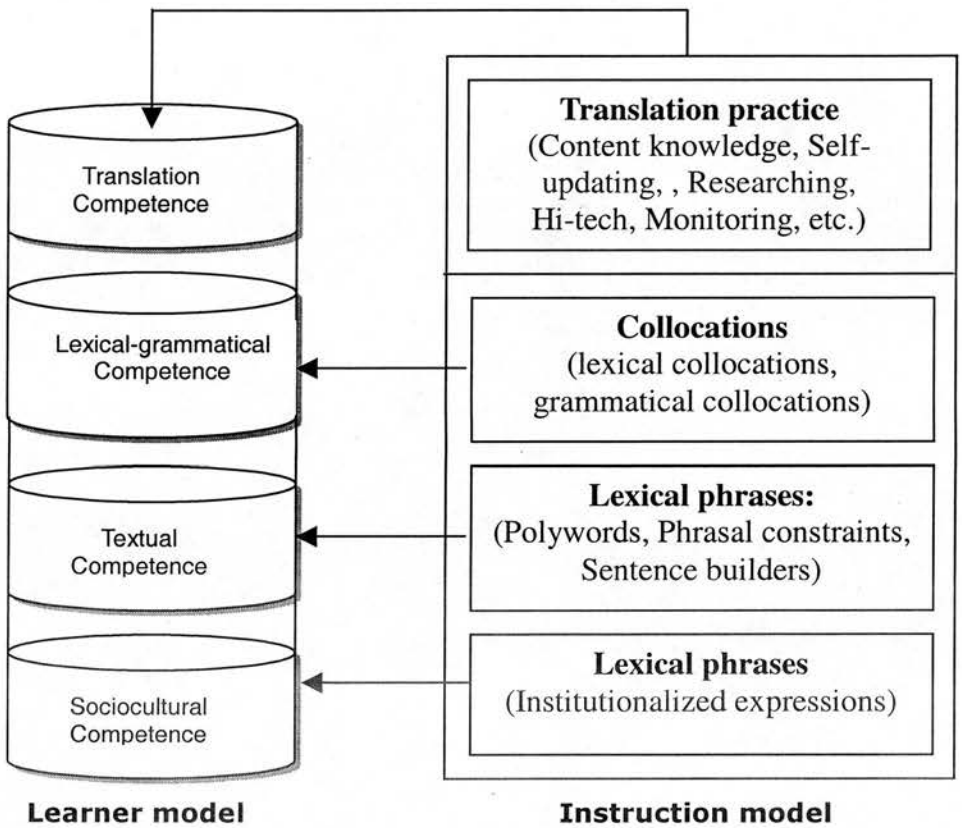


Figure 5-1: A model for teaching translation into the second language

This model highlights the idiomatic aspect of linguistic competence and language teaching, but it does not exclude the learning of grammar and vocabulary. Instead, grammar and vocabulary are placed in the background and the acquisition of both is understood to be a 'routine', since the learning materials and opportunities for both are abundant and learners should not need special prompting or help to access them. The objective for a student translator should not be writing grammatical sentences in the target language only, however. The ultimate goal for a translator should always be to render highly readable and idiomatic target texts. In order for the translation

trainees to acquire the ability to write idiomatic target language, the language learning module in a translation program should emphasise knowledge and ability in collocations and lexical phrases of the target language.

In Figure 5-1, there are three modules of instruction which aim specifically at the three linguistic modules in the learner model respectively:

1. Collocations: These include both lexical collocations such as *painfully aware* and grammatical collocations like *be busy with* (See 4.1.2). Adequate knowledge in English collocations enables the translation learner to tackle the issue of idiomaticity at the lexical-grammatical level.
2. Lexical phrases at the textual level: These include Nattinger & DeCarrico's Polywords (e.g. *in a nutshell*), Phrasal constraints (e.g. *as far as I* ___), and Sentence builders (e.g. *not only X but also Y*). A good command in these phrasal categories helps the translator with the task of achieving a good discourse structure.
3. Lexical phrases at the sociocultural level: These include Nattinger & DeCarrico's Institutionalized expressions (e.g. *the public seldom forgives twice*). Internalising a good repertoire of these and using them appropriately gives extra strength to the target texts by evoking rich sociocultural implications on the part of the reader.

The above listing leaves us with the explanation of the translation teaching module in the instruction model. As discussed earlier in Chapter 3, the teaching of translation has not attracted a large body of research on theories and pedagogies in history. Besides the prevailing on-the-job training approach, translation teaching has largely been conducted following the 'contrastive analysis' methodology, which however was recently challenged by the 'translator as communicator' view.

A good modern translator should be knowledgeable in many aspects of human life (Content knowledge), constantly in a state of upgrading one's skills and knowledge related to language and translation (Self-updating), knowing how to research and using reference materials (Research), always catching up with the latest high-tech developments (High-tech), and knowing how to correctly monitor one's progress and edit one's translation drafts (Monitoring). Note from Figure 5-1 that the translation teaching module has no specific contents, as it involves all aspects of translational, linguistic, sociocultural and communicational training.

The relationship between the translation module and the student model as a whole can be seen to resemble the spiral movement in language acquisition described by Brown (1996) (See 3.1.1) where linguistic competence is constantly being updated

by performance factors, both of which progress through interaction and refinement. In the case of our model, the translation module in the instruction model is the performance factor, which constantly informs and modifies the student's competence both in language and in translation. The student's competence in both language and translation as such, will in turn direct the translation activity module towards higher performance, which again upgrades the student module, and so on and so forth. This is, in fact, the theoretical foundation upon which the various implementation models described in the next chapters are based.

5.5 Summary

This chapter takes the view that instance-based processing is the primary language processing mode for human beings. Therefore, the learning of conventionalised forms and collocations, which are essential for this mode of processing, are essential for successful second language learners. A questionnaire investigation is reported in this chapter to verify the author's view that, for Chinese learners of English, knowledge in English collocations is an inadequate area which needs strengthening. A model for teaching translation into L2 is then proposed which emphasises the acquisition of collocations and conventionalised forms at the linguistic level. It is argued that the practice of translation is a good performance factor which can be used to test the learner's hypothesis in the idiomatic aspects of the target language. While translation serves to test and refine the learner's competence in L2, the advancement in linguistic competence, on the other hand, helps to shape the learner's performance, i.e. to produce better translation.

Chapter 6

Computers, Translation, Language Learning

This chapter discusses the role a computer plays in translating and in language learning. It examines the concept of 'translator workstation', explaining the use of concordancers, translation memory software, and machine translation techniques in translation. It also briefly introduces computer assisted language learning, focusing on a recent trend of incorporating corpora in intelligent tutoring. The chapter hopes to show that translation and language learning can converge felicitously in a computer aided instruction framework, especially that involving the use of corpora.

In terms of text manipulation, computers offer the facilities of large storage, quick search and retrieval and networked communication. This strongly influences the styles and efficiency of text processing, including those related to language learning and translation. The collection of large electronic corpora offers examples and generalisations of language usages, which benefit language teaching and help produce high-quality translation. The electronic tools like dictionaries, concordancers, encyclopaedias and Web page search engines provide instant solutions to inquiries or puzzles about word or phrase usages. The ever-improving technology of machine translation becomes increasingly helpful not only to business and industry but also for translators and language learners. The related technology of machine-aided translation also has important bearings both on translation and on language learning. Computer programs in various forms which combine characteristics of human intelligence in offering guidance with machine powers in handling resources and tasks, promise to become good assistants for translators and language users and learners in the future. The compelling need for pedagogists in language and translation teaching to address the issues of course orientation in the new computer era cannot be neglected. This chapter takes the reader on a brief tour from the translator's workstation to the translation memory applications, the machine translation technology and computer assisted language learning. The intention is to reveal the close relationship between translating, language learning and the role the computer can play in these settings.

6.1 Translator workstation

In modern times, a translator usually does not have to physically go to the library to find reference materials: a networked computer can provide answers to most

questions arising in translating. The reference tools residing in the computer or on the Internet include: electronic dictionaries, terminology banks, encyclopaedias, concordancers, collocation finders, Web search engines, etc. This computerised working environment has been referred to as the translator's workstation (Melby, 1992; Hutchins, 1997; Freigang, 1998) or the translator's workbench (Trujillo, 1999).

Melby (1992) distinguishes three levels of functions for a translator workstation. Level one functions include word processing, terminology management and telecommunications. Level two functions include text analysis, automatic dictionary look-up, and synchronized bilingual text retrieval. Level three "provides an interface to machine translation systems".

The use of electronic dictionaries, encyclopaedias and bilingual terminology files at the lower levels is straightforward and need not be discussed here. One helpful tool for the translator at level one which is less obvious to the novice translator may be the Web search engine. For example, in attempting to translate the text in (1) into Chinese, the Chinese equivalent for *Teletubbies* may not be found in the dictionary or terminology bank. The translator may know what the word means but it is another thing to translate it into Chinese, for an established translation may exist which the translator is not aware of.

(1) The characters from the cult children's television show the Teletubbies may seem unlikely classroom aids, but research shows they can help small children develop literacy skills.

In cases like this it is useful to resort to a localised Web search engine like Yahoo! Taiwan, which can help retrieve relevant local Web pages containing the accepted translations for the desired term. In this case, the Chinese term *tian-xian bao-bao* ('antennae babies') can be found to be the official translation for *Teletubbies* from some related Taiwanese Web pages retrieved by Web search engines.

Melby's (1992) level two functions include text analysis which incorporates mainly the tool of concordancers. It should be pointed out here that a concordancer is useful to the translator primarily in the target language domain. That is, it mainly helps the translator in the processes of rendering target texts (as opposed to understanding the source texts). The concordancer can search a large native-speaker corpus and find lines of texts containing the query word whose usage the translator, as a non-native speaker, is unsure about. The concordance lines provide the translator with a set of contexts that the target word is used in so that she can be more certain about how to integrate the word in her own sentence. For example, suppose our task is to translate the Chinese sentence in (2).

(2)

yi tade leguan , yiding keyi kefu nanguan
BASED-ON HIS OPTIMISM CERTAINLY CAN OVERCOME BLOCKADE
"Based on his optimism, he will surely overcome the blockade"

Since the literal translation for the first part of the sentence 'based on his optimism' does not seem to go very well with the rest of the sentence, the translator probably wants to look up a concordancer supported by a large corpus to find out how *optimism* is usually used by native speakers. The output from a concordancer could look like that in (3).

(3)

In May, there had been some	optimism that the four factions were
the linkage. Is there any	optimism that this international conference
senergy industry were full of	optimism that George Bush would help
any effort to bolster Jean's innate	optimism. The sight of me, unshaven and
With an unusual show of	optimism, the dour-faced Lutzemberger
had caused him to abandon Marxist	optimism. The crisis simply went beyond
he also had the great courage and	optimism (together with a jaunty cockney
Parton, 1986, p. 516). The `rule of	optimism" was seen by the report to be
comfort her and encourage her. Her	optimism was infectious. Joni's outlook
has been replaced by boundless	optimism. `We've done very well out of
something on which to base their	optimism when the reputable Washington-
developments will restore your	optimism, you'll regain control of your
Full of youthful	optimism and his consuming interest in

Note *full of optimism* appears twice in (3) and suggests itself to the translator as a likely phrase to use in the target sentence. Now consider the second half of the source sentence in (2). The translator can and should wonder if the literal translation of *overcome the blockade* is an acceptable collocation in English. The translator does well to query the corpus through a concordancer, this time using a 'collocation finder', which can list all the words that commonly co-occur with the query word in the corpus. If the translator would like to keep *overcome* and replace *blockade* with another word, the translator should enter *overcome* to look for its collocates. The results could be like that in (4).

(4)[34]

Collocate	Corpus Freq	Joint Freq	Significance
to	1375856	874	20.498440
can	133764	118	8.464217
problems	14436	77	8.454519
be	293540	165	8.394021
problem	15689	68	7.875621
by	230325	134	7.700213
help	25410	67	7.580680
difficulties	2135	50	7.012256
had	189551	104	6.577591
their	138506	80	5.927958
has	168986	88	5.872001
must	26913	41	5.584427
have	266301	110	5.542370
obstacles	503	28	5.272987

The translator can then decide that 'overcome problems/difficulties/obstacles' are all good collocations in English (from their high *t*-scores), among which the translator probably wants to choose *obstacles* for its stronger similarity to *blockade*. Thus the Chinese sentence in (2) is translation to a more idiomatic English sentence in (5) with the help of concordancers.

(5)

yi tade leguan , yiding keyi kefu nanguan

"Full of optimism, he will surely overcome the obstacles"

The use of concordancers in language learning has been amply suggested by various researchers and practitioners (Stevens, 1995; Gavioli, 1997; Johns, 1997; Cobb; 1999, Whistle; 1999). Most agree on the value of concordancers in helping the learner 'discover' rules and facts about language. Some point out, however, that students should not be left alone (helplessly) with a concordancer, but should be carefully guided in using it. The truth is: the use of a concordancer is not intuitively clear to a learner, its advantages not immediately obvious. On the other hand, it can be seen that if the language learner happens to be a trainee translator as described above, the situation is different. The use of a concordancer in this case is goal-oriented, and the procedures for finding desired information about language are

[34] Source: CobuildDirect Collocation Sampler at http://titania.cobuild.collins.co.uk/form.html.

clear.

Another level-two function in Melby's scheme of a translator workstation is what he terms the 'synchronized retrieval' of bilingual text files. A commonly accepted contemporary term for this is the 'translation memory', which is much more complicated than the foregoing passive look-up facilities and warrants a separate discussion.

6.2 Translation memory

Hutchins (1997) describes the translation memory as a facility which "enables the storage of and access to existing translations for later (partial) reuse or revision or as sources of example translations". According to Hutchins, "the sales of translator workstations incorporating translation memories are increasing rapidly, particularly in Europe". In Asia, some translation companies also advertise the use of translation memory software as an advantage on their Web pages. The practical value of translation memory software to translators is attested by its world-wide adoption by international enterprises, freelance translators and translation companies.

A piece of modern translation memory software, for example the TRADOS Translator's Workbench, usually has the following modules:

- Translation memory
- Terminology management
- Sentence alignment

The translation memory module is responsible for recording the user's translations sentence by sentence in a translation memory file. It also has a fuzzy-match algorithm which searches the memory to find any (partial) match for a new source sentence and retrieve the corresponding target sentence for the translator to consider. The terminology management module, on the other hand, keeps a bilingual terminology bank and, upon the activation of a source sentence, searches and retrieves the target language parts of any terms that exist both in the source sentence and in the database. Finally, the sentence alignment module allows the translator to import an existing piece of translation (from a source text file and a corresponding target text file) that is not previously treated by the translation memory module, and align the source texts with the target text on a sentence-to-sentence basis. The results of the alignment can then be exported to a translation memory file to add to its contents.

This amounts to saying that the whole team of translators can pool all their translation texts and terms together, including those from the present and the past, in

the form of translation memory files and terminology database files. This makes the tasks of translation and maintenance more efficient, the translation of terminology and recurring patterns more consistent across persons and organisations.

Though it is not immediately clear how translation memory can be related to second language learning, there is at least one case involving the use of translation memory for foreign language learning in a translation setting. DeCesaris (1996) reports on a translation program which uses translation memory to let Catalan-speaking students acquire grammatical knowledge in English. This works by giving students assignments of translating Catalan texts to English and allowing them to see the pre-installed English translations in the translation memory afterwards in order to learn the carefully arranged grammatical points embedded in the source text.

In theoretical terms, translation memory applications belong to the "machine-aided translation" category and are distinct from "machine translation" proper. However, both Hutchins (1997) and Freigang (1998) predict that the new generation translation memory software will incorporate a machine translation module so that the system will provide a translation anyway for any source sentence which cannot find a match in the translation memory. Thus it is time to discuss machine translation in relation to human translators and language learners.

6.3 Machine translation

Machine translation (MT) has had a fluctuating history since its origin in the 1940's (see Hutchins, 1995; Somers, 1998b). The fluctuations occur in terms of the methodology of MT: from the direct translation method to the transfer method, the interlingua method, to the more recent corpus-based method (see Somers, 1998b, Arnold et al., 1994; Trujillo, 1999); in terms of the expectations of MT: from FAHQMT (fully automatic high-quality MT), to no prospect at all in the late 1960's, to the current limited and realistic goals and expectations (see Hutchins, 1999); and also in terms of the platforms and users for MT: from the translation companies' desktops to the real-time Internet service for the general public (Yang & Lange, 1998) to its possible roles in the full-blown "teletranslation industry" envisaged by O'Hagan (1996).

As observed by researchers in MT, the success of MT is contingent on the restrictions of the domain, the input, the readers and purposes, and the language pair used (Hutchins, 1995, 1999; Somers, 1998a; O'Hagan, 1996: 26-35; Manning & Schütze, 1999: 463). It should be no surprise to readers that the MT research between Chinese and English is still in its youth and its current practical value is even

more restricted, since MT research started in the West and linguistic and computational linguistic research in Chinese is much less mature than that in Western languages. The Golden Bridge Translation Center is one of the leading Chinese machine translation research and application institutes, which offers a Web interface for machine translation between Chinese and English.[35] An informal test of its Chinese-to-English translation ability shows the clause-level translation to be quite good; whereas its performance at the discourse level has still room for improvement, as the test sentences in (6) show.[36]

(6)

a. Single clause

yi tade leguan , yiding keyi kefu nanguan
BASED-ON HIS OPTIMISM CERTAINLY CAN OVERCOME BLOCKADE
"Can overcome the difficulty certainly with his optimism"

b. Multiple clauses

ziji de xinyang rang ta dui xuduo ren de piping
SELF 'S BELIEF LET HIM TOWARD MANY PEOPLE 'S CRITICISM
bu gandao shengqi , bing xiangxin zhu hui baohu ta ,
NOT FEEL ANGRY , AND BELIEVE LORD WILL PROTECT HIM ,
zhiyao yu zhu tongzai , jiu juede hen anxin .
SIMPLY WITH LORD TOGETHER THEN FEEL VERY REASSURED .
"One's own faith let between he and a lot of criticism of people feel angry , and believe that mainly protect him, So long as with coexist mainly , think and settle down very much right away."

We can see that the MT's output in (6a) is almost right except that the target sentence is short of a subject (*he*), which is characteristic of many Chinese sentences. We can also see MT failing miserably in (6b) when it tries to analyse multiple-clause Chinese text. This seems to indicate that human translators will still play a crucial role in the English-to-Chinese translation (or Chinese-to-English translation, which is equally hard) market for a long time to come.

Though MT in the future will play a more and more important role in the language services of the World, it will not replace human translators in the foreseeable future. As Hutchins (1998) points out,

---

[35] This is at: http://www.netat.net/english/index.htm.

[36] Transcription convention: first line: Chinese Pinyin system, second line: word-for-word translation, third line: machine translation

> ...it will become common knowledge that automatic translation cannot, unaided, produce anything better than 'rough' (occasionally barely comprehensible) 'translations' ... It will be recognised that for higher quality, e.g. for translations of publishable quality, the solution will remain with human translators, aided by all the computer aids that are appropriate.

The same observation is made by O'Hagan (1996: 35), who says that "users are recognising the potential benefits of MT and are beginning to make use of the technology within restricted environments". In other words, MT "must no longer be put forward as a 'solution' to people's translation needs, but it must be seen as no more than a 'useful aid'" (Hutchins, 1999). Increasingly, not only is MT used by the general public for dissemination (of information), communication, entertainment and language learning purposes (Yang & Lange, 1998) in this restricted sense, it is also used by translation professionals, as per O'Hagan's (1996: 32) comment that "the 10 leading translation companies in Japan had each implemented some form of MT".

Not surprisingly, students soon find that MT can help them "do their foreign language homework" (Yang & Lange, 1998). For a language teacher, on the other hand, MT can also become a useful teaching aid. A good example of applying MT to language learning is Anderson (1995), who reports on a Hebrew-to-English machine translation system used for English speakers to learn Hebrew. Because the MT system produces imperfect English sentences, the learner is asked to check a reference bilingual corpus and other tools to find out why the MT system would make such a mistake. The more sophisticated MT technology becomes, the more possibilities of integrating MT in a language teaching curriculum.

6.4 CALL, ICALL, ITS, corpora

Computer assisted language learning (CALL) does not have a long history, since the computer itself is a relatively recent invention in human history. However, as computer software and hardware industries move forward, possible ways of applying new technologies to CALL also increase infinitely. A more recent variant of CALL is ICALL (intelligent CALL), which is a kind of ITS (intelligent tutoring system) specifically used for language learning. The integration of corpora into CALL is the latest significant development of CALL and is directly relevant to the application dimension of this thesis.

Higgins (1995: 4-6) makes a figurative distinction between four types of CALL programs:

- DO WHAT I TELL YOU: The computer decides the nature and order of events.
- GUESS WHAT WAS THERE: The student guesses words and restores texts 'hidden' by the computer.
- CAN I HELP YOU?: Students explore the language with the computer as a tool.
- HOW DO I GET OUT OF THIS?: This consists of activities such as games and simulations.

Kenning (1990: 51-78) discusses categories of CALL from the aspect of classroom procedures (presentation, practice, and production), isolating CALL programs such as drills, vocabulary practice, text reconstruction, concordance packages, simulating conversation, drawing packages, adventure games, problem-solving, etc.

ICALL is a later development of CALL with the addition of elements drawn from the field of artificial intelligence. A typical ICALL program will contain some of the modules discussed by Lian (1992):

- A domain expert
- A student model
- A teaching expert
- A natural language interface
- A reasoning system

Similarly, Burns & Capps (1988) consider three knowledge foundations of an ITS -- "expert knowledge, student diagnostic knowledge, and the instructional or curricular knowledge", which correspond to Lian's first three modules. Thus, the ICALL software should be 'more intelligent' than the previous CALL systems in terms of the machine's domain knowledge, and its ability to monitor students' progress and to carry out teaching instructions.

Of the various advantages of CALL recognised by scholars, one which is especially relevant to this thesis, is the ability of the computer to offer a large quantity of authentic material for learning and its capacity for encouraging authentic use of language. Kenning (1990: 80) comments, "With regard to communicative activities, the value of the computer stems from its ability to help create a context favourable to authentic uses of language." Also, among Chanier's (1994) nine proposed requirements for ICALL, one is that "ICALL must offer rich, authentic language input", and the other is that ICALL must use "interesting and relevant themes" and "meaningful language tasks" to allow the student to learn the materials "through

associations".[37] Pennington (1996) further points out that

> ...as a consequence of the amount and variety of types of input made accessible by the machine, students often increase their risk-taking behavior and experimentation with learning and with language.

If this is true, authentic-text based CALL programs with some kind of student modelling module, such as the one described in Shei (2001), which introduces an 'automatic language lesson generation' system that can create a language learning unit based on any text the student chooses and can monitor the user's progress in terms of his progress in vocabulary of the foreign language, seem to be a good example of materialising essence of both CALL and ICALL. More generally, a recent development of using corpora for language teaching in a computerised environment is actually making the most of CALL in an unprecedented way, in terms of its ability to exploit authentic language usage.[38] A typical CALL program integrated with corpora is reported in Johns (1997), whose aim is "to exclude invented examples of English as far as possible, whether for presentation or for practice" and the materials used "are based on examples of authentic usage recovered from a corpus by means of a concordancer".

6.5 Summing it up

Reflecting on the above-discussed educational and translational technology, we seem to see a many-way convergence in terms of computers, translation and language learning. We see translation being combined with computers in the implementation of a translator's workstation. We see language learning being integrated with a computer system in the CALL environment. Furthermore, the possible combination of translation (especially translating into L2) and language learning is also strongly hinted at from the general working environment they share (i.e. both benefit greatly from the use of computers), the tools and resources they use (e.g. concordancers and corpora) and their common goal of interaction (i.e. to produce high-quality written language approaching native speakers' standard). Their relationships are as Figure 6-1 shows.

---

[37] The reader will be reminded here that doing translation meets exactly these two criteria for ICALL in Chanier's scheme: translation exercises usually consist of authentic texts, and doing translation is easily perceived by students as performing an authentic task as compared with doing a language structure exercise.

[38] See Murison-Bowie 1996 for a review of using corpora in language teaching.

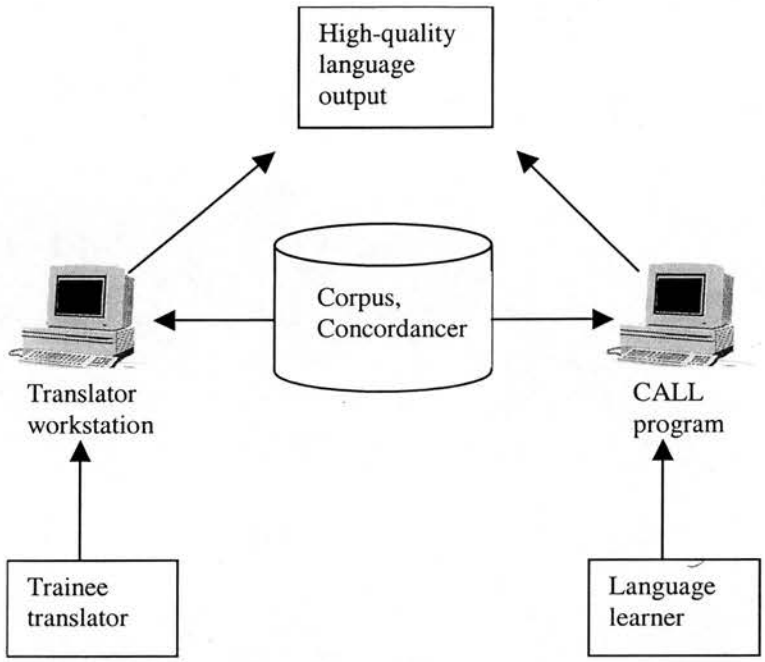


Figure 6-1: The convergence of translation and language learning in a computerised environment

It seems reasonable to suggest, then, that translating into L2 and L2 learning activities can be conducted in an integrated computerised environment and both benefit doubly as a result of mutual reinforcement and other factors such as expanded resources and interest.

Looking back at the model of instruction we proposed at Figure 5-1, where translation practice is used to enhance the trainee translator's idiomatic abilities, we seem to find it closely related to the situation represented in Figure 6-1, i.e. the recent trends in the practice of translation and in language teaching using CALL. In other words, the idea of integrating translator training with language learning seems very much in tune with where the translation and language learning technologies are heading. What is particularly worth noting is that the strengthening of the ability of translators in using collocations and lexical phrases, emphasised in Figure 5-1, projects readily onto one of the areas where the current CALL possibilities are most fruitfully exploited -- the use of concordancers and similar tools to explore corpora.

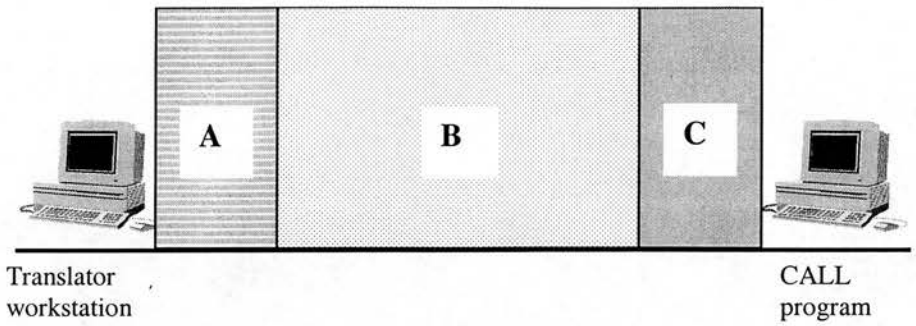


Figure 6-2: The application continuum between TWS and CALL

If we place the translation workstation (TWS) at the one end and general CALL programs at the other end, we have a continuum of computer applications from the job market use of TWS to the use of CALL in the educational setting, as represented in Figure 6-2. We can further distinguish three rough categories along this continuum between the two extremes:

A. TWS-based translation training programs: The most typical case of this category is using commercially available translation software *as is* in the translator training program with additional help from the instructor, such as the one reported by Talbot (1996). This could also be what Zanettin (1998) calls "Translator Trainee Workstation", which puts together a word processor and commercially available bilingual corpora and concordancing facilities to aid in the training of translators. Or what Wältermann (1994) terms "Pedagogical Translators Work-Station", which is transformed from an in-house developed, UNIX-based TWS with additional support on top of the regular TWS features. Programs which embed translation training materials in existing machine translation software (Anderson 1995) or translation memory software (DeCesaris, 1996) also belong to this category.

B. Middle ground: There is an indefinite range of possibilities for this category, each embracing different proportions of TW and CALL elements. One essential feature which distinguishes this category from A (emphasising the learning of translation skills) or C (emphasising the learning of language skills) would be the emphasis on simultaneous acquisition of language and translation skills. Software developed along this line would be most suitable for students learning to translate into a second language who need simultaneous help from both ends. The Translation Micro Worlds to be introduced in Chapter Eight will fall into this category.

C. CALL programs involving translation: These would be CALL programs which

use translation as a means for teaching language skills. Since the Grammar-Translation method has long fallen into public disfavour, research reports exploiting translation as an instruction method are rarely seen, not to mention CALL programs. Duff (1989) presents one of the rare examples of teaching language through especially designed translation exercises, which if adapted into a CALL program would be an example of this category.

In summary, while recognising the value of classroom instruction by human teachers in the traditional way, we should not neglect the tremendous influence of computer technology on translation and language education. Most of all, this area of intersection between translation and language learning, with the computer in the background, looks promising to translation educators who are seeking a solution for training student translators learning to translate into the second language, as well as to L2 practitioners who are interested in using translation as a vehicle for L2 learning.

Chapter 7

Translation Software Based Systems

We proposed a continuum of computer software for learning language to translation in the last chapter (Figure 6-2). In this and the following two chapters, we will look at each of the three types of learning software between the translator workstation and general CALL programs. This chapter first describes the TWS-based translation training programs, i.e. the Type A applications in Figure 6-2, in which the data to be processed by the translation software is arranged in a pedagogically meaningful way, such that the student using the system is not only learning to use the software, but also, learning the pedagogical points intended by the course instructor. We first focus on how a piece of translation memory software can be used for teaching students concepts or skills in translation. We then look at the possibility for incorporating machine translation software in a translation curriculum.

7.1 Introducing TRADOS

TRADOS is one of the mainstream commercial translation memory systems. (referenced by DeCesaris, 1996; Hutchins, 1999). It works by remembering every pair of source-target sentences the translator has created, and displays any matched pair of stored sentences when activated by a new source sentence. Thus, it is claimed, the translator never has to translate the same sentence a second time. Through a fuzzy matching facility the translator can also retrieve translations of source sentences similar to the current source sentence and edit them instead of translating from scratch. The workflow is summarised in Figure 7-1.

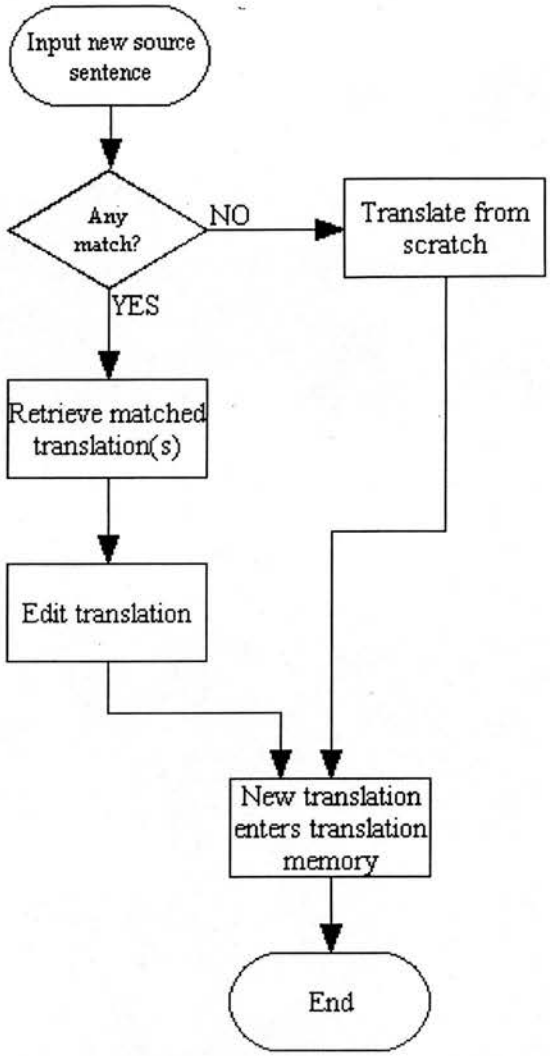


Figure 7-1: How the TRADOS translation memory works

To investigate how we can make TRADOS work for a course in translation, we need to have a look at its application interfaces as well. TRADOS 5, the edition we use in CJU, has a set of five tools, whose names and functions are as follows:

- Translator's Workbench: This is the main application which creates, updates and searches the translation memory file and retrieves the matches.
- MultiTerm: This is a terminology database which works closely with the Translator's Workbench, providing translation for any matched terms in the source sentence. The Translator's Workbench matches whole sentences drawing from the translation memory; whereas MultiTerm matches single terms from the separately constructed terminology bank.
- TagEditor: This is an application which lets the translator work on html files using TRADOS's standard translation tools.
- T-Window for PowerPoint: As the name implies, this is TRADOS's tool for

working with PowerPoint files. This runs as a plug-in program and the translator works in the standard Microsoft PowerPoint environment plus TRADOS's translational tools which live within the application.

- WinAlign: This program allows the translator to import a pair of source text and target text files and align the texts bilingually sentence by sentence. These bilingually aligned sentences can then be imported into a translation memory file as further assets for providing translation matches in later translation tasks.

In the following sections, an overview is offered for each component listed above.

7.2 Translation Memory

The translation memory module includes a memory file, search facilities and abilities to communicate with other applications such as the translation editor and the term bank. In TRADOS, the application which manages these tasks is called the Translator's Workbench.

7.2.1 Translation Memory File

The translation memory file is a type of file unique to each kind of TMS software, which usually includes aligned bilingual translation units and format information, etc. The translation memory file in TRADOS, when opened in the Workbench, consists of bilingual texts like the ones illustrated in Figure 7-2:

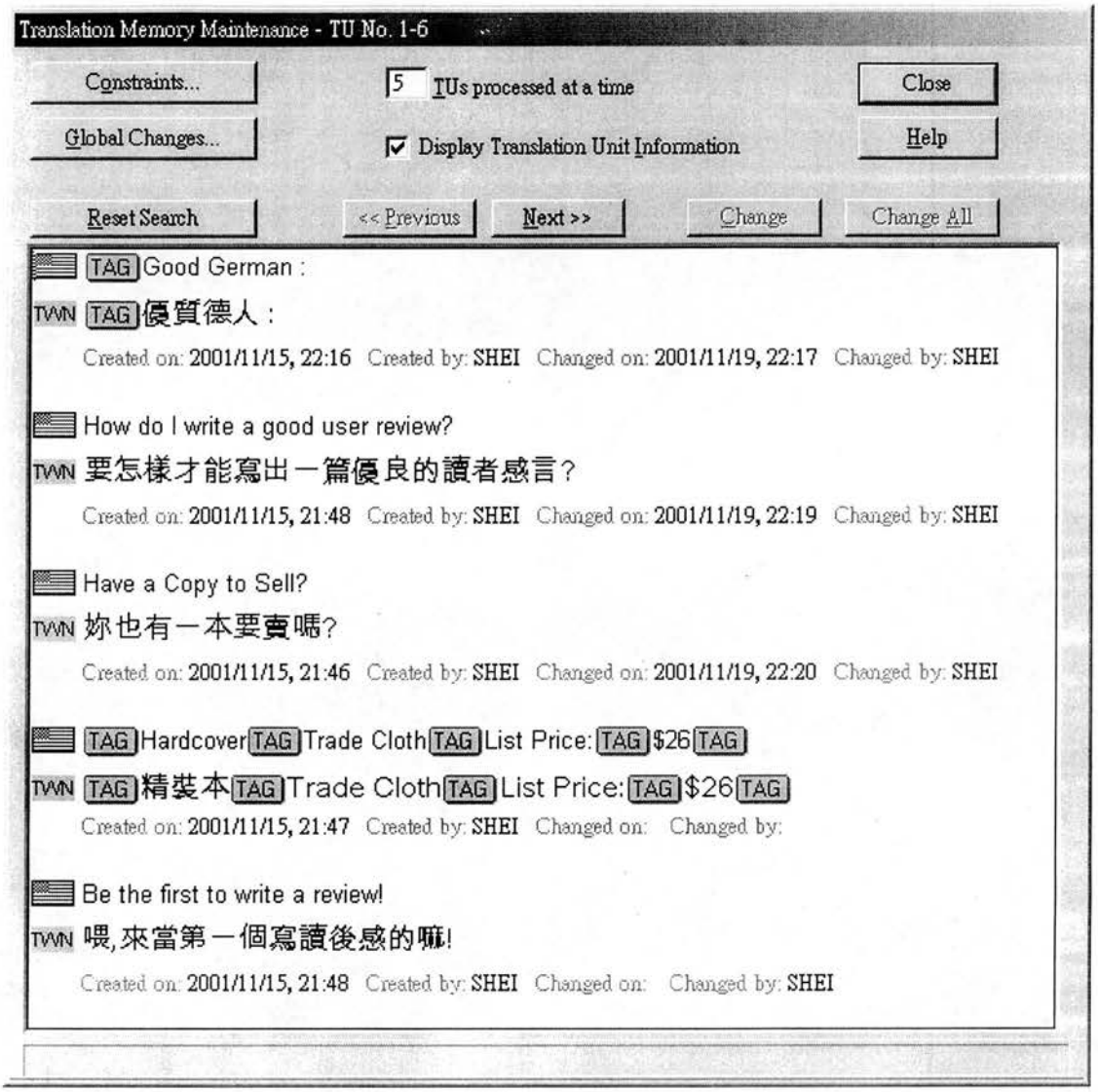


Figure 7-2: An example content of a translation memory file

7.2.2 Search and Suggestion facility

When the user opens a new translation unit (often a sentence) to translate, the TMS runs the sentence through the translation memory and retrieves any bilingual pair(s) with identical or similar source sentence(s). In TRADOS Workbench, the result of a match looks like the one shown in Figure 7-3:

Figure 7-3: The example of a match between memory and input

The upper box in the Workbench interface of Figure 7-3 is the current source sentence being opened in the TagEditor – a kind of word-processing application especially created for translating documents with tags like the hypertext documents. The lower box, on the other hand, displays the bilingual pair existing in the translation memory which matches the source sentence in the upper box. Note the match value is 61%.

7.2.3 Communication facility

We have seen in Figure 7-3 that the main program of a TMS can communicate with a translation editor like the TagEditor. The other module the main program is connected to is the term base, from which the program can retrieve translations of terms found in the source text. For example, the TRADOS Workbench contains another window for displaying the term matches, as shown in Figure 7-4:

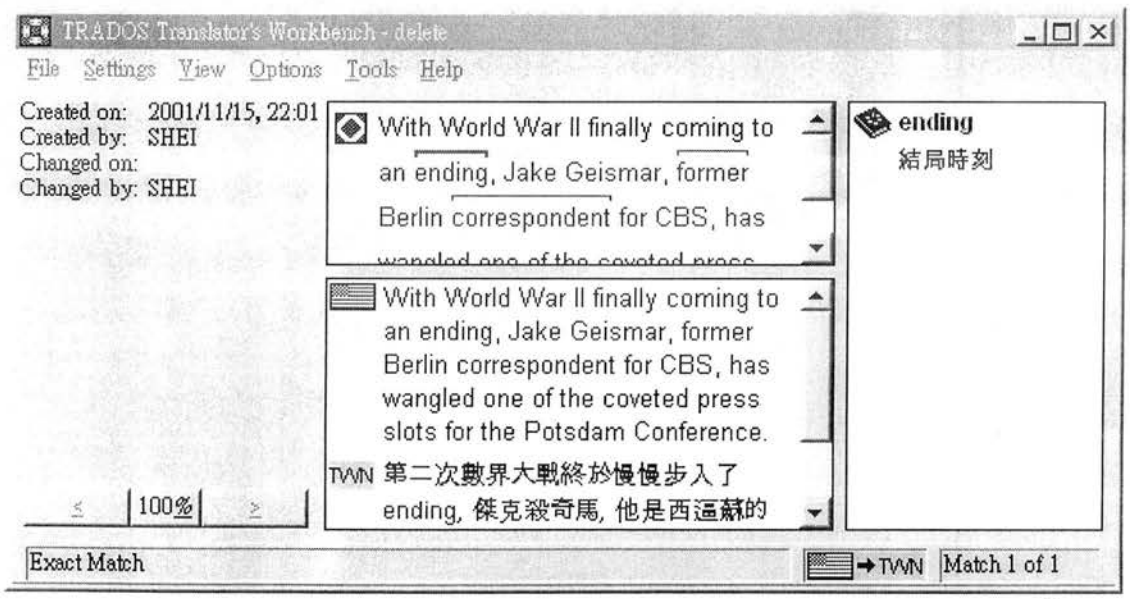


Figure 7-4: TRADOS Workbench retrieving terms from the term bank

If the user finds the translation of the term useful, he can import it directly into the appropriate position of the target sentence in the translation editor.

7.2.4 Basic Machinery

According to Trujillo (1999), the translation memory system uses a search mechanism much like that use in Information Retrieval (IR). Here the query is the newly entered source sentence for translation, and the database (DB) is the translation memory (TM). The basic principle is to find a sentence in the TM DB which has the highest number of the same words as the input sentence. In order to avoid favouring longer sentences in finding a match, however, the similarity counts (numbers of identical words between two sentences) need to be normalised using a formula like Dice's coefficient (Trujillo, 1999: 62):

$$M = \frac{2|I \text{ I } T|}{|I| + |T|}$$

Here M is the similarity measure, I is the set of words in the input sentence and T is the set of words in the translation unit (TU) being considered. Suppose for the sentence *How to write a good user review*, we want to find out which of (1a) and (1b) is more similar to the input:

(1)

a. A user asked the author of the review how he could claim that the book was good.
b. What does a good user review look like?

If we simply count the same words, then (1a) has four and (1b) has only three words which also appear in *How to write a good user review*, though in fact (1b) seems to be intuitively more similar to the input sentence. Now if we use Dice's coefficient, we get:

For (1a): $2\times\frac{4}{7+17}=0.33$

For (1b): $2\times\frac{3}{7+8}=0.4$

So in fact (1b) is the more similar sentence rather than (1a), since its *M* score is higher (i.e. a higher proportion of words match).

According to Trujillo (1999), the translation memory system makes use of a kind of index file to facilitate the search procedure. The algorithm first makes an index file (also known as the 'inverted file') of all the words in the TM DB along with the codes of the TUs where they appear. When encountering a new input sentence, the algorithm finds all the words appearing in the index that match the input words, and calculates the similarity measure. Those translation units with the highest scores will be retrieved. For example, suppose there is a mini TM consisting of the few sentences in (2):[39]

(2)

a. Books message boards.
b. What does a good user review look like?
c. Be the first to write a user review!

Before attempting to build an index file, the algorithm does a 'cleaning up' job first by stemming the words (getting rid of the inflectional suffixes), and taking out words in the 'stoplist'. The stoplist consists of the most frequent words like *the, a, of, to*, etc. the inclusion of which would make the results of calculation unusable, as this greatly increases the chance of any sentence matching any other simply due to the frequent appearance of these function words. Thus after stemming and applying the stoplist we get the index file in (3) for the database in (2):

[39] There will be translations for these sentences too in the actual translation memory, but the target sentences are omitted here.

(3)

Words in TM	Host TU
be	c
board	a
book	a
do	b
first	c
good	b
like	b
look	b
message	a
review	b, c
user	b, c
what	b
write	c

Given the above index file for the existing translation memory database, suppose the newly input sentence is *How do I write a good user review*? We calculate the similarity measure for each sentence in TM DB (2) to find out which is the most similar to the input sentence:

For (2a): $2\times\frac{0}{7+3}=0$

For (2b): $2\times\frac{4}{7+7}=0.57$

For (2c): $2\times\frac{3}{7+5}=0.5$

Sentence (2b) has the highest similarity measure, the TU consisting of (2b) and its translation will be displayed (e.g. in the lower box of Figure 2) as the best match to the source sentence *How do I write a good user review*. As for the other matches, the TMS usually allows the user to define minimum match value (e.g. 60% similarity) and the maximal number of hits.

7.3 Translation Editor

In contrast to the idea of constructing a 'universal translation editor' which can let the

translator work on files of all formats,[40] TRADOS's solution is to introduce its own translational tools in the form of plug-in programs into existing industrial-standard editing programs. For example, when the Translator's Workbench is installed, a Trados menu will appear in the menu bar of Microsoft Word and a new tool bar will also appear populated with translational tools which work interactively with Translator's Workbench within Word. Figure 7-5 illustrates the co-operation between Word and Translator's Workbench when doing a Chinese-to-English translation task.

Figure 7-5: Word cooperating with Translator's Workbench

Figure 7-5 shows an Aesop fable being translated from Chinese to English. (For convenience of illustration, the windows of Translator's Workbench and Word have

[40] This is indeed an approach another major kind of TM software, Transit, takes.

been put one on top of the other.) Note the two (originally coloured) boxes in Word which contain texts: the top box accommodates the source sentence (Chinese) currently being worked on; whereas the lower box displays the match in translation memory provided by the Workbench, if any. Texts above these two boxes are target texts already translated. Texts below are source sentences to be consecutively translated. Note also the *Trados* menu item and the extra translational tool bar added by TRADOS in the Word window, which are used to communicate data between Word and Translator's Workbench -- i.e. to retrieve existing translations from translation memory and put new source-target pairs in storage. In the Workbench itself, on the other hand, there are two main windows and one side window. The two main windows contain the current source sentence and its analysis (top) and the source-target pair retrieved from the translation memory (down). As for the side window, this is used to display terms from the source sentence with their translation equivalents retrieved by the MultiTerm application, which can be conveniently imported into the document opened for translation in Word.

For other types of documents, for example, for translating Web pages, TRADOS develops a separate interface called TagEditor as the translation workspace. Figure 7-6 shows an example interface:

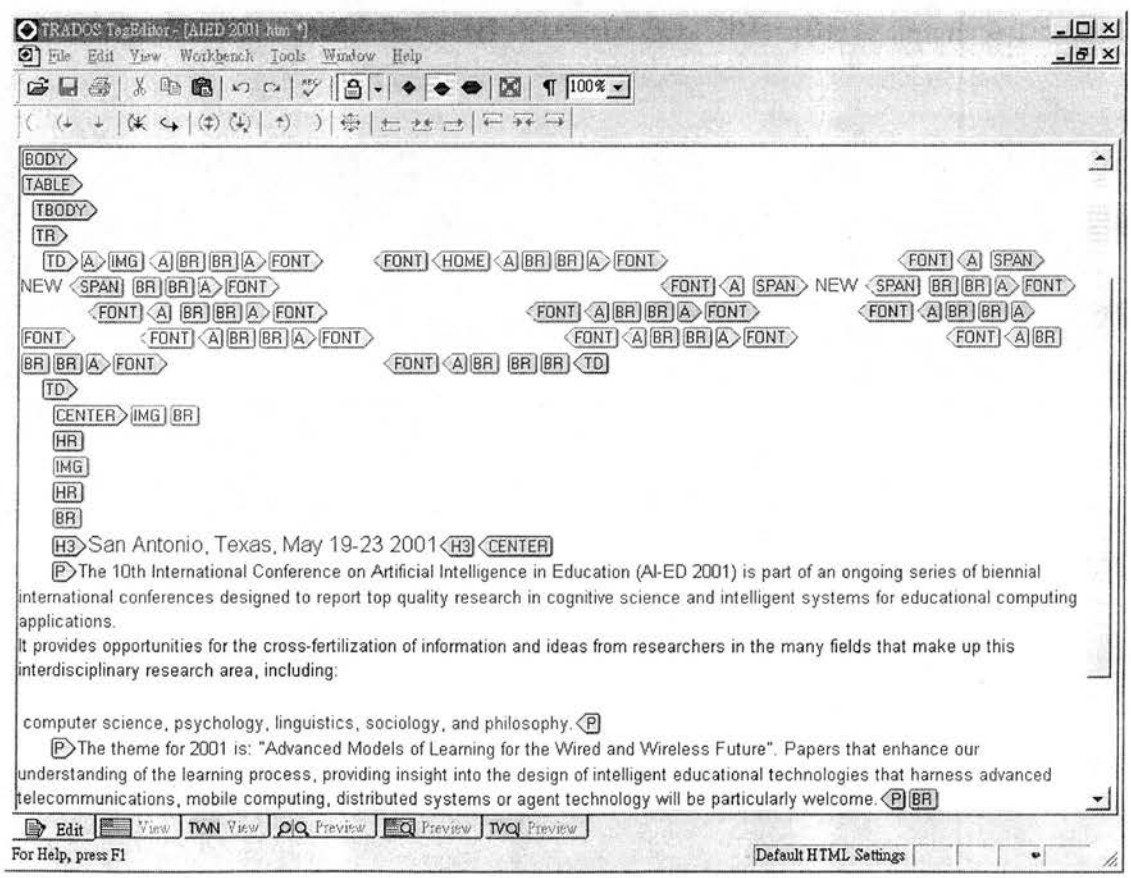


Figure 7-6: A TU opened in TagEditor

Note the document is filled with HTML tags which must be kept in order for the translated Web page to be in exactly the same format as the original, including the hyperlinks, etc. Note also that there is a TRADOS tool bar in TagEditor as well, consisting of various combinations of arcs and arrowed lines, the same as the one shown to live with Word in Figure 7-5, which is used to communicate with Workbench.

There is also a T-Window interface especially designed for translating Microsoft PowerPoint files, which works on the same principle.

7.4 Term Management

The TMS usually includes a terminology component which is like a bilingual (or multilingual) dictionary. When a translation unit is being opened in the translation environment, an algorithm searches the term database automatically in the background and calls up any identical or similar terms with their translations for consideration. In TRADOS the terminology management module is called MultiTerm, which is shown to be communicating with Workbench while working on a TU in Figure 7-7:

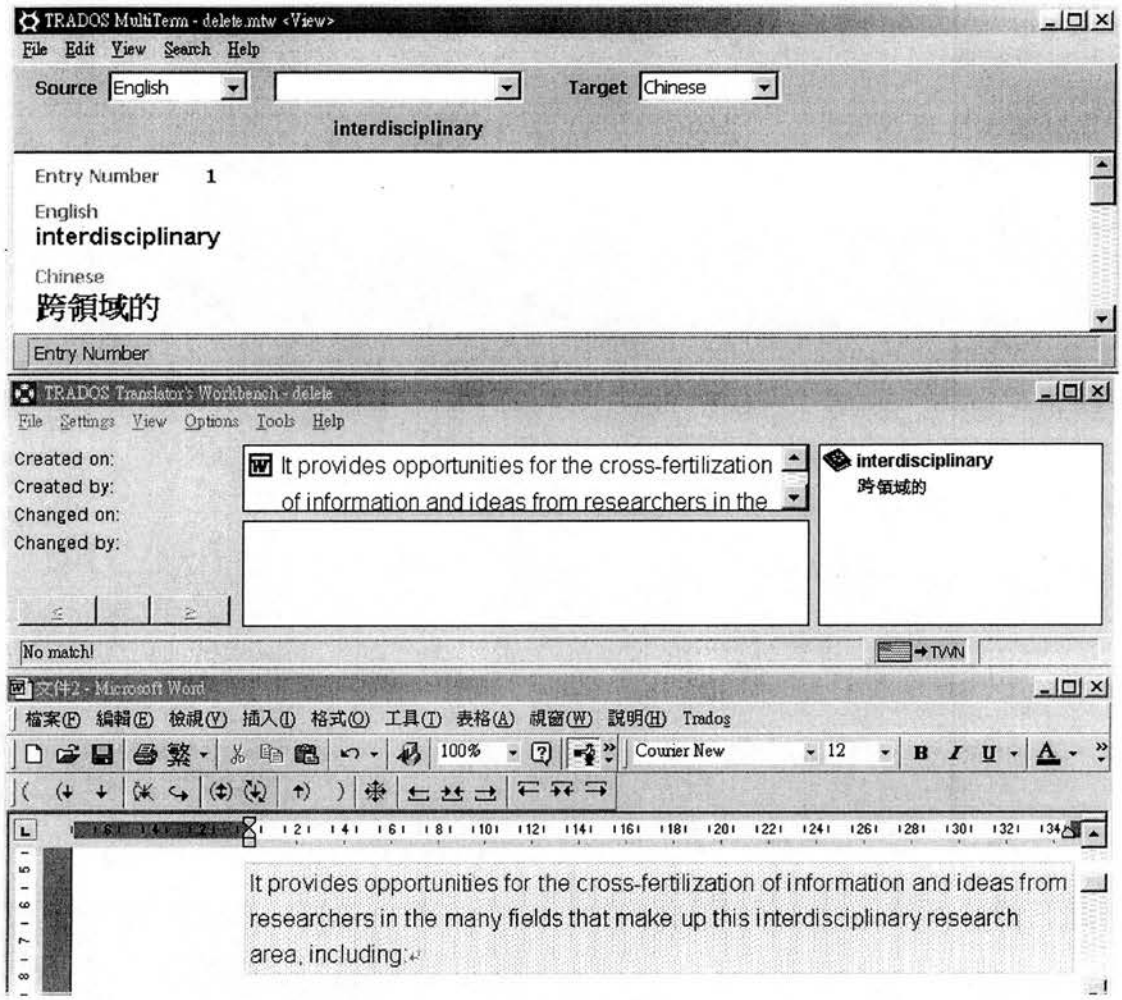


Figure 7-7: A TU in Word getting terminology support from MultiTerm

In Figure 7-7, a translation unit in the source language (English) is opened for processing in Word. The same sentence also appears in Workbench, where a separate column shows a term retrievable from MultiTerm. The same term itself also shows up in MultiTerm together with any relevant information.

7.5 Alignment

A full-fledged translation memory system like TRADOS also contains an application for aligning bilingual texts. This usually works on two files, one containing the source text and the other containing the corresponding target texts. The objective is to align the two files in such a way that each source sentence is integrated with the corresponding target sentence (i.e. its translation) to be stored in the TM as a translation unit. The purpose for doing so is to incorporate more TUs into the translation memory to increase the translator's resources. In TRADOS, such a program is called WinAlign, as Figure 7-8 shows.

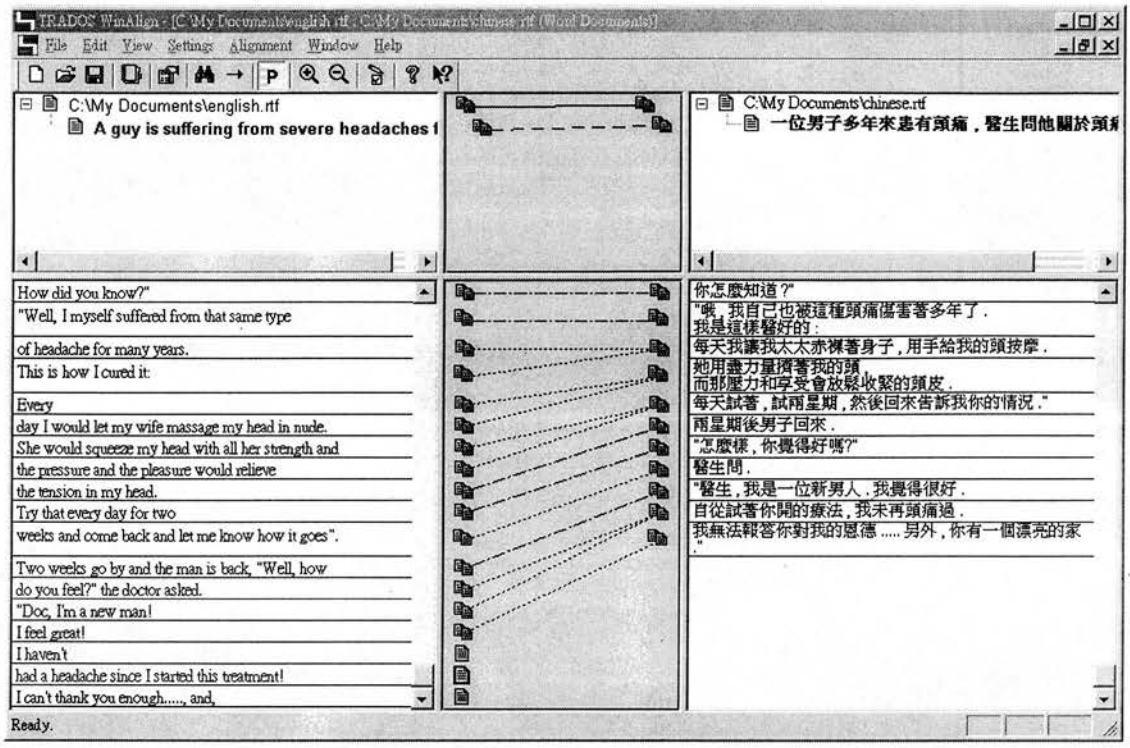


Figure 7-8: An alignment project conducted in WinAlign

Automatic alignment of bilingual texts by computer programs is possible and has attracted a number of researchers in the field of computational linguistics. However, to guarantee 100% accuracy as is required by the professional standard, some human intervention is necessary in the alignment process. In the case of TRADOS WinAlign, automatic alignment is provided in the first instance (as Figure 7-8 shows). Manual adjustment by the translator must then follow to achieve the ideal result of alignment before it can be exported to a translation memory.

7.6 Using TRADOS for teaching translation

Before continuing with this topic, a distinction must be made between teaching students to use TRADOS and using TRADOS to teach students other things (e.g. a specific translation technique). This section is concerned with the latter; that is, it attempts to show how the inherent facilities of TRADOS can be used to host some types of translational exercises to teach students concepts or skills in translating.

DeCesaris (1996) briefly hints how "a controlled model" can be embedded in the translation memory, and then how translation tasks can be designed which "make students query the translation memory so that they access the translation model that the teacher wants them to emulate". DeCesaris' example involves the translation of

an authentic Catalan text into English, where a standard English translation is preinstalled in the translation memory as the correct model. The source text given to the student for translating is an altered version of the original Catalan text with grammatical points embedded in the alterations which are meant to draw the student's attention to potential problems for native speakers of Catalan writing English. Not enough details are given in DeCesaris (1996) about how the system works, but the idea is clear and it represents what is indicated in Figure 6-2 as a Type A program; that is, a program which is based on a piece of software typically included in a translator workstation, which is used in a pedagogical rather than professional way.

Because of the 'retrieve and edit' working style of TRADOS, it is not difficult to design a grammatical kind of exercise, where the target sentences kept in the translation memory can be intentionally and incorrectly altered in some way (depending on what grammatical points are being focused on) or some critical words or structures can be left out altogether. When the student translates the source text sentence by sentence, each corresponding target sentence with some grammatical flaws will be called out for the student to edit. After finishing translating the student can compare her version with the standard version and learn from her mistakes.

For example, if the current lesson in translation focuses on the selection of English verbs while translating from Chinese into English, and the source text to be worked on is the Chinese translation of the Aesop fable *The Hare and the Tortoise*, the standard English version of the story is shown in (4) with the Chinese version following as (5).

(4)

A HARE one day ridiculed the short feet and slow pace of the Tortoise, who replied, laughing: "Though you be swift as the wind, I will beat you in a race." The Hare, believing her assertion to be simply impossible, assented to the proposal; and they agreed that the Fox should choose the course and fix the goal. On the day appointed for the race the two started together. The Tortoise never for a moment stopped, but went on with a slow but steady pace straight to the end of the course. The Hare, lying down by the wayside, fell fast asleep. At last waking up, and moving as fast as he could, he saw the Tortoise had reached the goal, and was comfortably dozing after her fatigue. Slow but steady wins the race.

(5)
有一天，一隻兔子嘲笑烏龜的腳短，走路慢。烏龜笑著說：「雖然你走得很快，像風一樣，假若我和你賽跑一次，我會贏你呢。」兔子認爲他的主張是不可能的，便依了他的提議；議定請狐狸選擇跑道，指定終點。到了賽跑的那一天，他們一同出發。烏龜一步也不停的，總是用慢而穩健的步伐，直向那跑道的終點前進。兔子在中途便躺了下去，酣睡起來。後來醒了，雖盡力的向前奔跑，烏龜卻早已到了終點。辛苦之後，正安適地在那裏微睡。緩慢而穩健就會贏得比賽。

While the standard English version in (4) will be kept as a correct model for reference, the instructor needs to modify the English text based on the desired learning effect. In this case, in order to let student ponder problems in the choice of English verbs, the easiest thing to do is of course to leave all the verbs out in the English part of the translation memory for the student translator to fill in. The teacher does this by preparing the translation in TRADOS himself and intentionally alters the target sentences so that the source sentences and their corresponding edited target sentences enter the translation memory in the desirable way. So for example, when the student activates the first sentence in (5) in order to translate, Word presents the interface as shown in Figure 7-9.

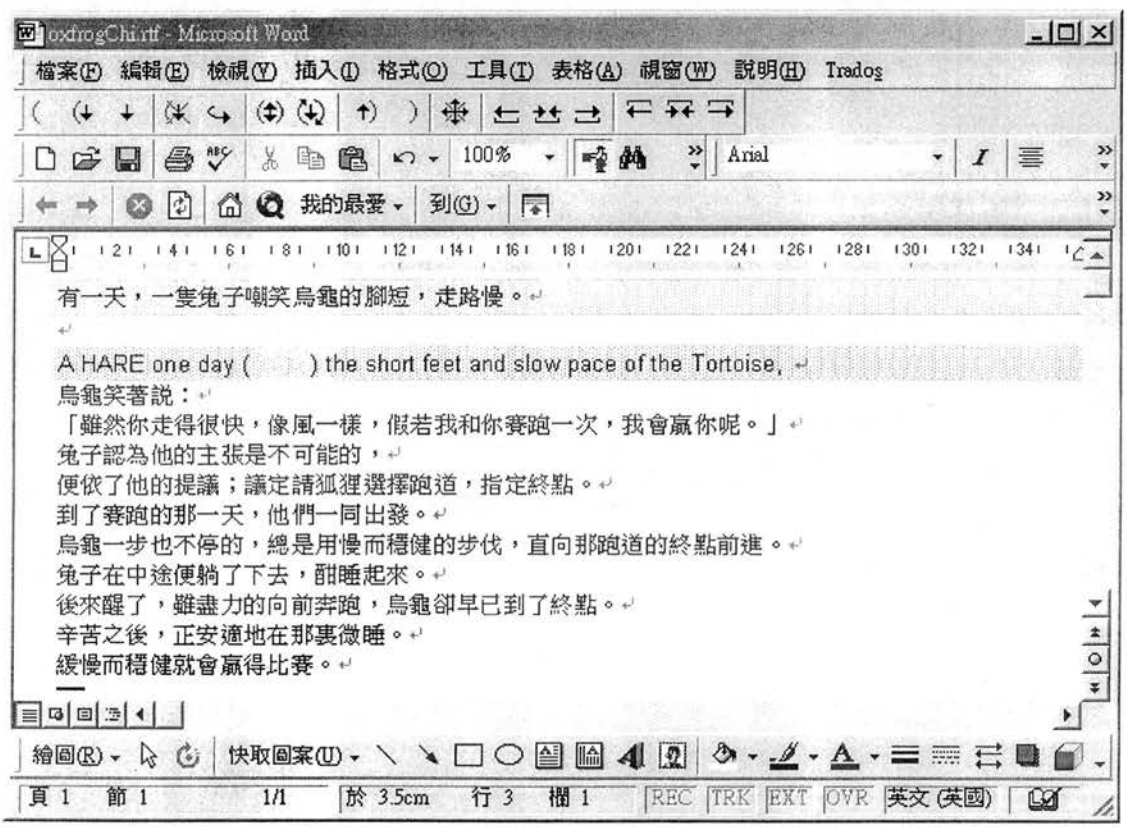


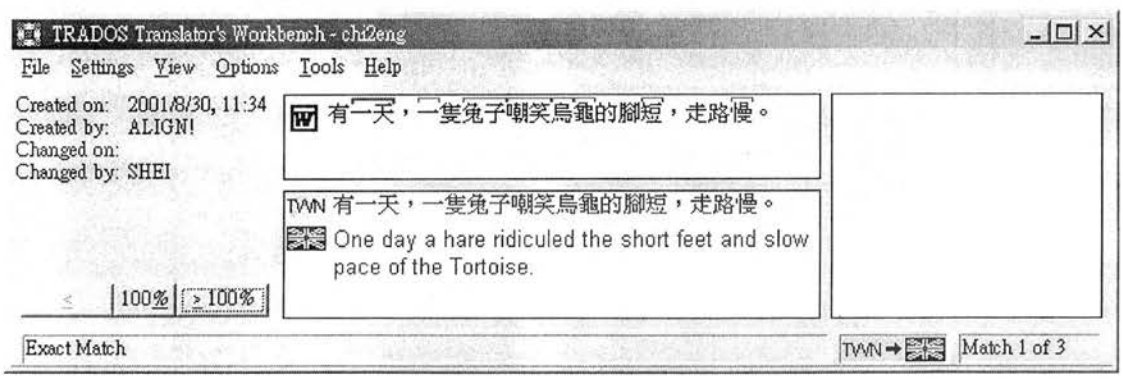
Figure 7-9: An embedded pedagogically modified English sentence

In this example the student compares the source sentence with the target sentence and tries to locate a verb to use based on the Chinese word and the context. Hints can also be provided, for example in the form of several synonyms to the verb originally used, which can be entered into TRADOS's MultiTerm and incorporated in the entry of the corresponding Chinese verb and be called out upon activation. After the student finishes translating the whole passage by supplying all the required verbs, she can be provided with the original English text and compare the usage of the verbs. A general profile of the student's ability to use English verbs can be obtained and any misconceptions may be clarified.

It is also possible to embed multiple matches in Translator's Workbench for a given source sentence. For example, for the first Chinese sentence in Figure 7-9, the instructor can prepare at least the three different translations shown in (6).

(6)
a. A hare *one day* ridiculed the short feet and slow pace of the Tortoise.
b. *One day* a hare ridiculed the short feet and slow pace of the Tortoise.
c. A hare ridiculed the short feet and slow pace of the Tortoise *one day*.

This can be done in a translation lesson focusing on the placement of English adjuncts. All the moveable adjuncts in (4), for example, can be moved about to form multiple matches for the corresponding source sentence in the translation memory. The student can move between these choices of target sentences and decide on the most satisfactory one to use. The way how the alternative sentences in (6) are displayed in Translator's Workbench is shown in Figure 7-10.



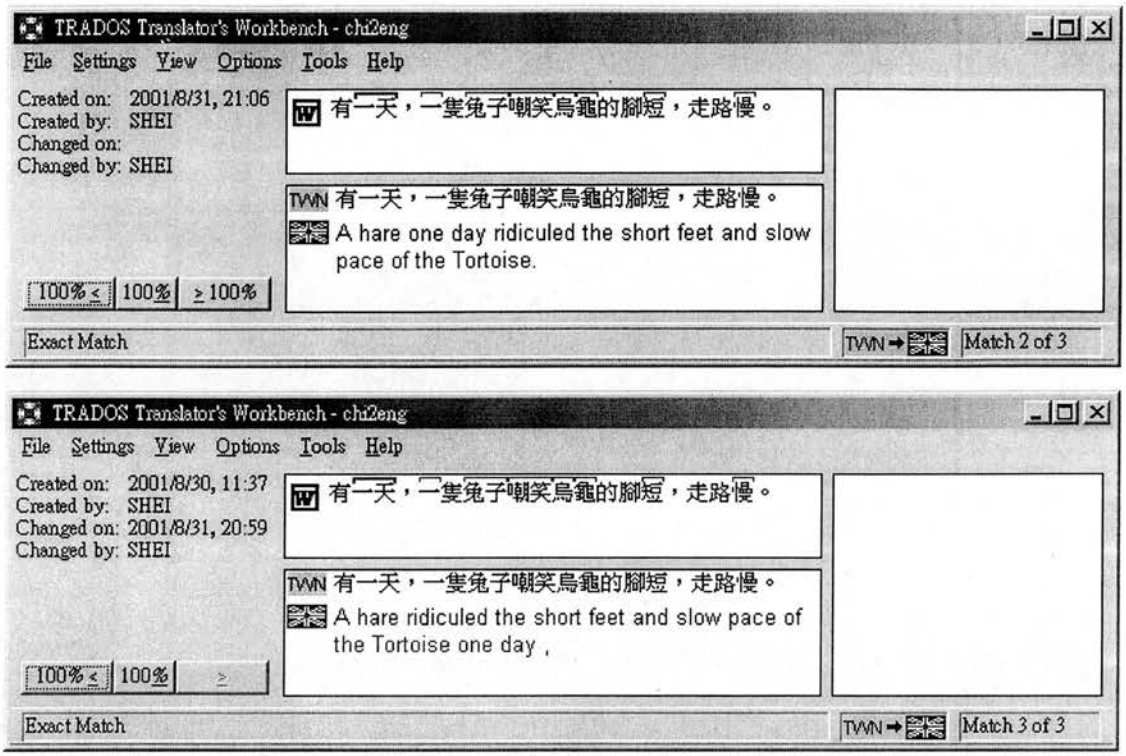


Figure 7-10: Three slightly different matches for the same source sentence

Because in Chinese it is not possible to move the adjunct *you yi tian* ('one day') freely within the sentence, instruction methods like this are valuable in that they let students appreciate the various ways of arranging adjuncts in a sentence and their different connotations.

Thus it is conceivable that translation memory systems can be used in a variety of ways following the manner introduced above. The trainee translators take these exercises not only to familiarise themselves with this tool which will play a crucial role in their future careers, but also to learn translational techniques or concepts especially prepared by the instructor.

7.7 Using MT for teaching translation

To date Anderson (1995) still seems to be a rare example of using machine translation system as a tool in language learning, to the present author's awareness. It is understandable since the output of the state of the art MT can hardly qualify itself as a good model for language learning. However, this does not preclude it from being used as a stimulant or as a means for the learner to emulate a good language model, as Anderson (1995) does (see 6.3).

As we do not have any machine translation software between Chinese and English in CJU, which is rarely seen on the software market anyway, we use the

online MT system of the Golden Bridge Translation Center (GBTC) to illustrate the basic idea of how it can be used for teaching translation. For the Chinese version of the Aesop fable *The Hare and the Tortoise* shown in (5), the GBTC MT machinery translates it into the English passage in (7):

(7)
One day, one rabbit laugh at tortoise foot short, it is slow to walk. The tortoise says smiling: ' though you walk very fast, like wind, if you and I race once, I will win you. ' the rabbit thinks that his opinion is impossible, depends on his proposition; Discuss and decide and ask the fox to choose the runway, appoint terminal point. By that day racing, they set out together. One of tortoise suspend even , use slow and sane paces is frank to advance to that terminal point of runway always. The rabbit is lying down midway , sleeps soundly. Having woken up later, although what did the best ran forward, the tortoise has already reached terminal point. After working hard, are sleeping a little there comfortably . Slow and sane will win the match .

One can see that there is a large room for improvement in the English text generated by a machine translation system. Thus the most obvious way for using the outcome of MT in (7) is to let the student practise editing it into more acceptable English, comparing it with the standard English version in (4). This exercise benefits the trainee translator in at least two ways: first, the student can get used to the translating style of MT and learn how to collaborate with MT to produce useable translation; second, the student can learn from MT's mistakes and understand the subtle differences between good and bad translations. Before MT can start to produce translations of much better quality, 'editing or analysing the bad translations of MT' seems to be the only conceivable way of using MT in language or translation learning.[41]

7.8 Conclusion

Type A programs on the TWA-to-CALL continuum are the easiest to implement in a translation curriculum in that they consist of existing translation software which has a distinct place in the professional setting of translation. The advantage of adopting this kind of program is that it can be closely connected to the reality of the profession of translation, whether in student perception or at the level of skill transfer. The disadvantage is that there is no possibility for modifying the existing software adopted so the scope of application is quite limited. As we move on to the CALL end of the continuum, more flexibility in designing the software is allowed. As a result, the

[41] Though some may worry about the adverse effect of learners' over exposure to such 'bad language'.

pedagogical elements will increase as the commercial flavour of translational software decreases. In addition, the learning of translational skills gradually gives way to the learning of language through translation. A piece of software residing in the middle ground which achieves a good balance between translational professionalism and L2 learning will probably be the kind of software most sought after, which will be the topic of Chapter Eight.

7.9 Summary

This chapter explored the concept of a translator's workstation and discussed its possible components, which include concordancers, translation memory software, and machine translation. Translation memory software is relatively unknown to the layman on translation but is of crucial importance to the profession. A large part of this chapter was thus devoted to explaining its designing principle, implementation, and application in translation teaching. The relevance of concordancing and machine translation to the teaching of translation was also discussed.

Chapter 8

Translation Micro Worlds

In the last chapter we discussed how existing translation software can be used to teach translation, which is in effect a usage that deviates from the software's original application domain (i.e. being the tool of a professional translator). This is what we termed a Type A program. In this and the next chapter we will look at software specially developed for the teaching of translation (Type B) and second language learning (Type C) respectively. In this chapter we discuss the Type B program which differs from a Type A program in that Type B is custom-made for translator training and so is more powerful in accommodating teaching ingredients. It is also farther away from the professionalism of the translator workstation and closer to the pedagogical side of translation and language learning.

8.1 Translation Micro Worlds

The Translation Micro World (TMW) proposed here is not the kind of micro world in the CALL tradition which usually include graphics, moveable objects and intelligent characters, and natural language dialogues (e.g. Murray, 1995). The Translation Micro World proposed here contains language data of a particular kind which allows the user to practise translating in a controlled environment focusing on a specific learning goal. A typical TMW may consist of a parallel corpus, some linguistic units extracted from the corpus, some rules for manipulating the knowledge, and some optional components like user modelling modules, and a user interface. For the first TMW discussed below, the corpus comprises English and Chinese bilingual jokes, the linguistic elements extracted from the corpus are collocations and lexical phrases, and the rules are those regarding the manipulation of this idiomatic knowledge in the system's tutoring activities.

8.2 Collocation and lexical phrases tutor

Recall that we argued in Chapter 5 about the importance of the second language learner's (as well as the translator's) ability to use English collocations and lexical phrases. To conform to the instructional framework proposed in Figure 5-1, we need to encourage student translators to learn the idiomatic usage of foreign languages while practising translating. The collocation and lexical phrase TMW introduced in

this section is designed exactly with this in mind, where the student is led to learn English collocations and lexical phrases through the actual task of translating.

8.2.1 The corpus

The corpus on which the system is based is some fifty jokes collected from the Internet comprising more than one thousand words. Jokes are chosen because:

1. There is plenty of un-copyrighted material of this kind on the Internet which is suitable for an experimental system like this;
2. The language in many jokes is quite compact and idiomatic and suitable for the present purpose;
3. Jokes cover a wide range of topics and so are in tune with the versatile nature of translating.

An example joke collected in our corpus is shown in (1).

(1) GOOD FIRST IMPRESSION?[42]

> A young doctor was just setting up his first office when his secretary told him there was a man to see him. The doctor wanted to make a good first impression by having the man think he was successful and very busy. He told his secretary to show the man in.
>
> At that moment, the doctor picked up the telephone and pretended to be having a conversation with a patient. The man waited until the "conversation" was over. Then, the doctor put the telephone down and asked, "Can I help you?"
>
> To which the man replied, "No, I'm just here to connect your telephone."

第一印象

一位年輕醫師, 剛剛開業, 正在整理中. 秘書說, 有一位男士要見他. 醫師想要給他一個好的第一印象, 讓男士覺得他是成功與忙碌的. 醫師要秘書帶他進來.

這時候醫師拿起電話對著病人說話. 男士等著, 一直到對話停止. 然後, 醫師將電話放下說 : "有事嗎?"

對此, 男士回答 : "我來此爲你安裝電話."

For those readers who understand Chinese it will be obvious that the English text in (1) is the source text while the Chinese version is a translation. The quality of the

[42] Note the Chinese translation of the joke follows the English version without any linguistic annotation as it is used here only to convey the impression that this is a bilingual text. Our focus in this article is on the original English text and we do not analyse the Chinese counterpart or their correspondence in any way.

Chinese translation is low presumably because it is only done by amateurish translators 'for fun' on the Internet. In our TMW system, however, we need the student to practise translating Chinese into English; hence we have to rewrite all the Chinese texts like that in (1) into more idiomatic Chinese to serve as source texts in our system, while maintaining a certain degree of meaning and structural correspondence between the new Chinese text and the original English text.

8.2.2 The analysis

Once the corpus of jokes reaches a usable size, we start extracting collocations and lexical phrases from it manually. For example, from the text in (1) we obtain the idiomatic units in (2).

(2) Idiomatic extraction from (1)

Collocations:

- Lexical collocations: *set up office, make impression, have conversation, connect telephone*
- Grammatical collocations: *tell + NP + NP, have + NP + VP, tell + NP + to + VP*

Lexical phrases:

- Polywords: *at that moment*
- Phrasal constraints: *to show ___ in*
- Sentence builders: *be just X when Y, there be X to see Y*
- Institutionalized expressions: *can I help you*

Thus all the discernable idiomatic expressions of these six categories in the corpus are extracted and consolidated into a data bank for the use of the TMW, which is described in the next section.

8.2.3 The modules

The Jokes TMW has five components: the control centre, the student modelling (or translation memory) module, the user interface, the in-built corpus and the extractions from corpus. The relationships between them are described in Figure 8-1.

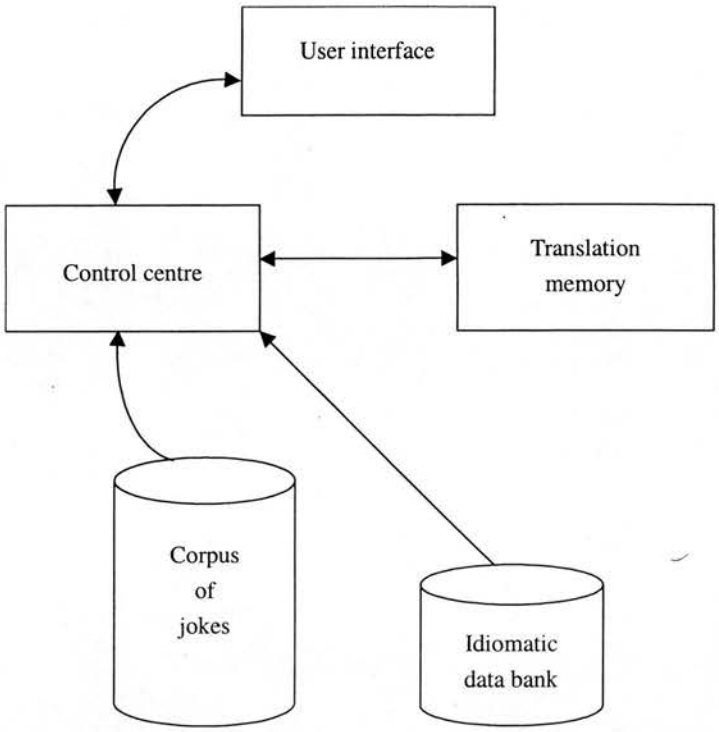


Figure 8-1: The structure of the Jokes Translation Micro World

- Control centre: It feeds jokes in Chinese to the user to practise translating into English. It provides the pre-stored idiomatic units associated with each source sentence upon the request of the translator. It updates the user profile in the student modelling module (i.e. the translation memory) in terms of the number and kinds of idiomatic chunks learned. Furthermore, it matches each Chinese sentence the user is currently working on against the user's personal translation memory bank to see if any idiomatic expressions can be retrieved for quick use.
- Student modelling module: It keeps in a file all the idiomatic chunks that the user has encountered while practising with the system and their translations in the student's mother tongue. The control centre can add to this file, as well as search and draw from this file, as in checking whether a particular idiomatic chunk has been met by the user before. This file can 'follow' the student across different TMWs, accumulating the mass of collocations and lexical phrases learned.
- User interface: This should be a stand-alone text editor or a few "macros" living within an existing word processor like Microsoft Word. It displays a unit of the corpus at a time and lets the user translate the text sentence by sentence. It provides the pre-stored translations for the user to see and to appreciate the idiomatic side of the language, to learn collocations and lexical phrases, and to have the chance of reviewing and reusing these linguistic devices.

8.2.4 An example session

The user first creates a translation memory file, which is a kind of database file. The user then imports a corpus, in this case the corpus of jokes, into the application interface. The user will notice a more unusual row of four buttons on the rightmost end of the tool bar in the application interface illustrated in Figure 1. Let's call them, from left to right, the *generator*, the *left arrow*, the *information*, and the *right arrow* button respectively. The user first presses the *generator* button to get the next joke to be translated. The translation is done sentence by sentence. The user highlights the sentence to be translated with the curser and press the *left arrow* button, which encircles the source sentence in a square box and produces another empty square box in which the user is to type in the target sentence, as Figure 8-2 shows.

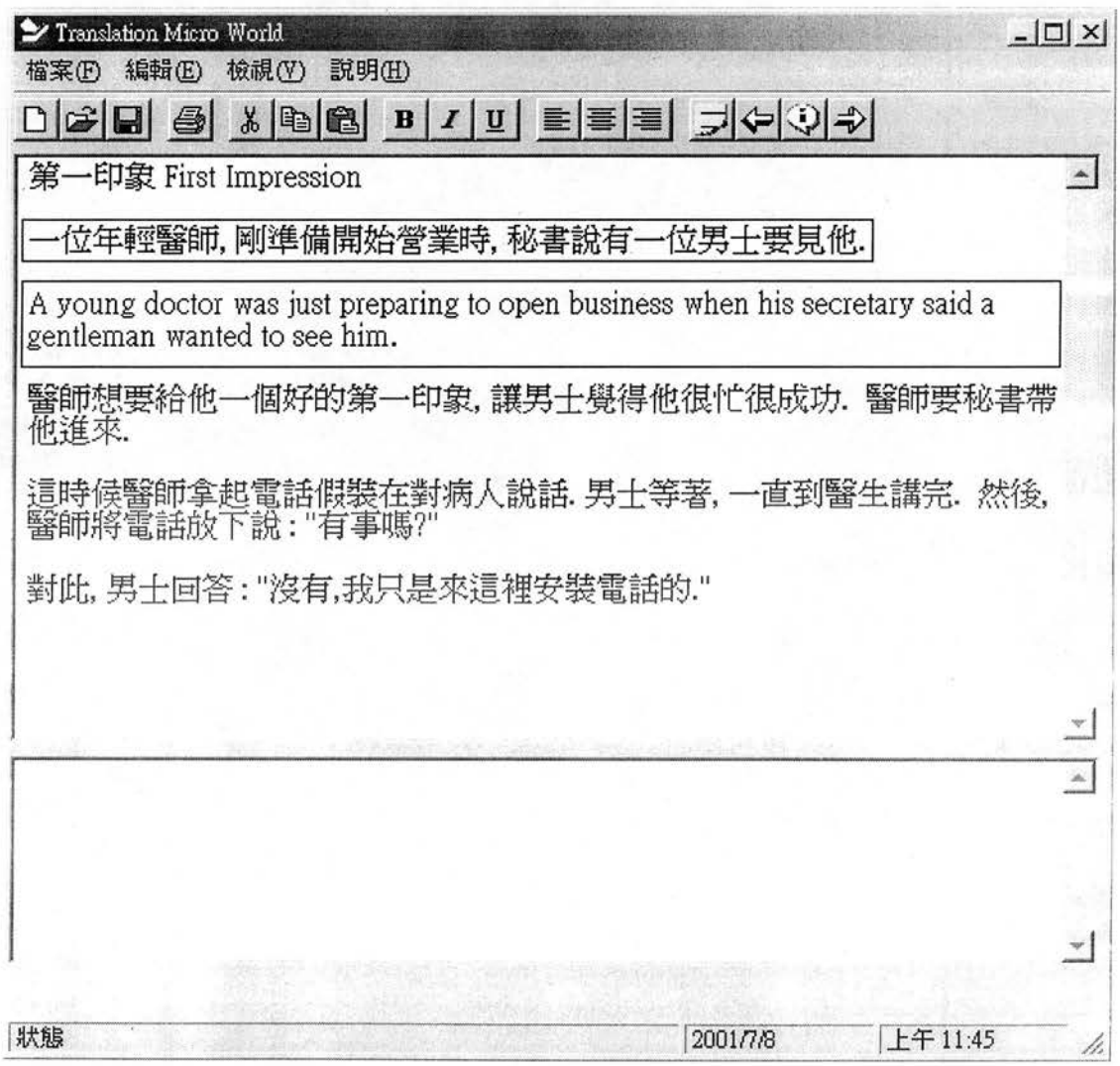


Figure 8-2. The Translation Micro World interface

After the user has done or considered the translation, she can click on the *information* button to view the idiomatic chunks or formulae related to the current sentence in the lower column of the program interface, as Figure 8-3 shows. The user can then modify her translation into more idiomatic language based on these hints.

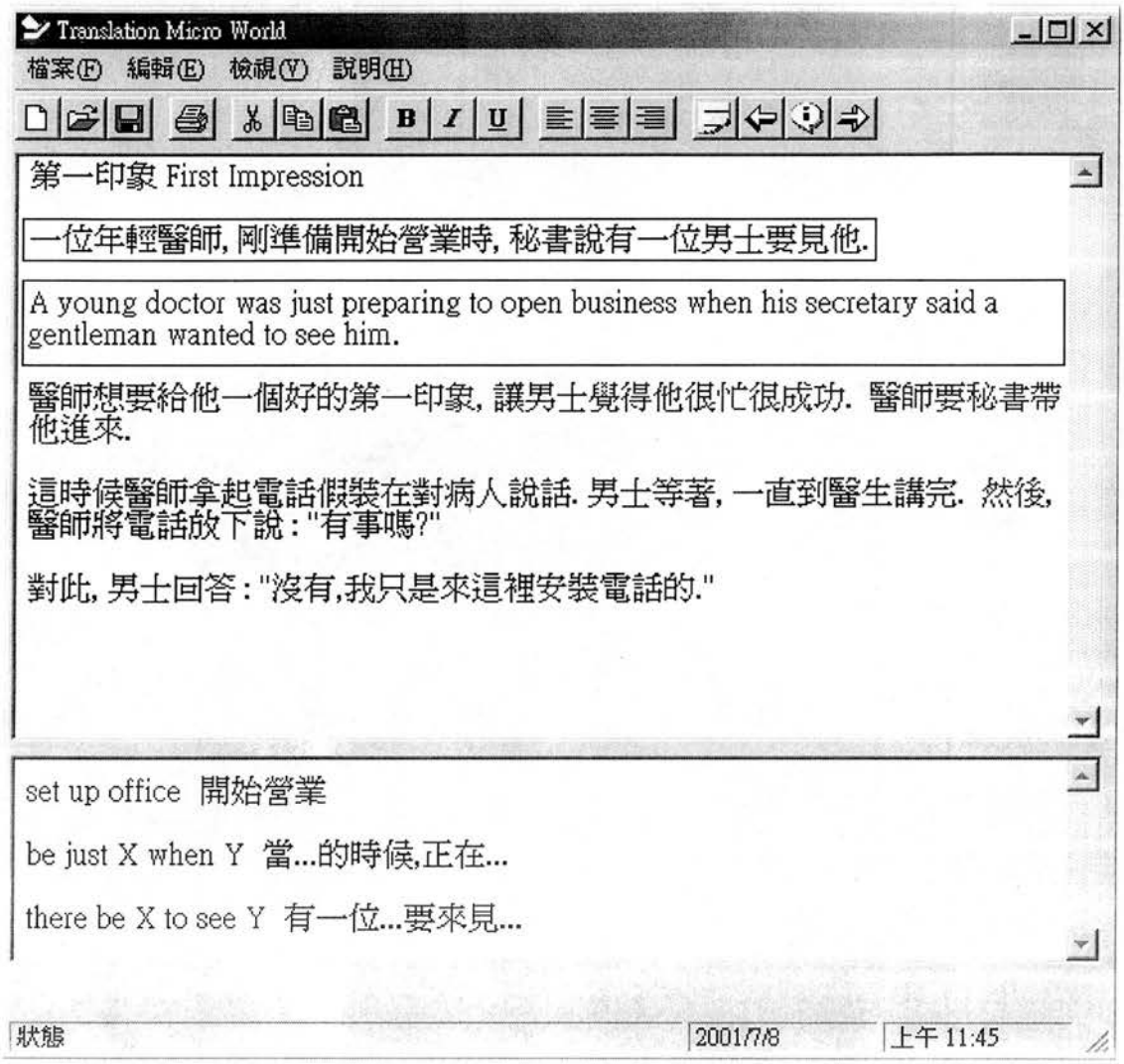


Figure 8-3. Idiomatic information offered in the lower textbox

Finally, the function of the *right arrow* is for the user to see the original English sentence stored in the system. The user can edit this sentence any way she likes, until she presses the *left arrow* again to highlight the next source sentence to be translated, as Figure 8-4 shows. Note the previously accomplished sentence has been edited by the user based on the system's suggestions of idiomatic usage.

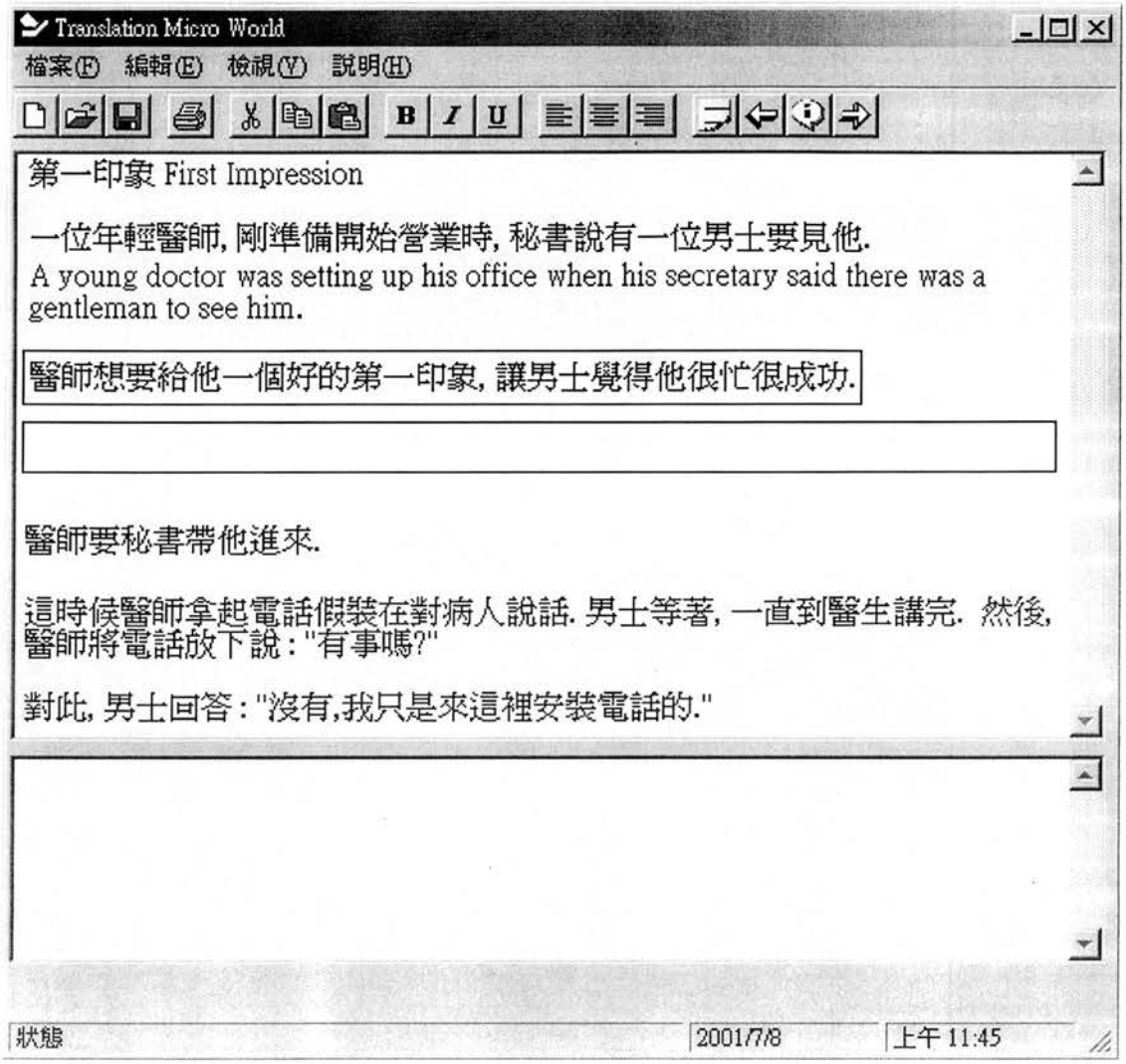


Figure 8-4. Moving on to the next source sentence

Thus, when the student finishes translating some jokes, she not only has the use of the entire translated texts, she also has a personal translation memory file which contains the bilingual versions of all the collocations and lexical phrases she has encountered and accepted in the practice session. This is further explained in the next section.

8.2.5 Translation memory

When the student starts using the TMW application, she is prompted to create a translation memory file, which is a database that is going to contain all the English collocations and lexical phrases with Chinese translations encountered and accepted by the student while using the system. The Chinese characters are later on used as indexes when the system searches for idiomatic matches in a given Chinese sentence for translating. The student will have accumulated a considerable repertoire of

collocations and lexical phrases of all kinds in her translation memory file after completing a course with a TMW. Henceforth, when the student is doing any work of translation in another TMW or in the real world on her own, she can always use the memory file with a translation editor which can communicate with this file, and be reminded of a particular collocation or lexical phrase when the corresponding Chinese text or its look-alikes are encountered. While the student does other learning activities such as reading in the foreign language or explicit grammar instructions to improve her grammar and vocabulary, the translation memory takes care of the idiomatic side of her second language acquisition by keeping a mental record for her and pulling out useful idiomatic expressions for her to use in translation.

8.2.6 Conclusion

The students at the department of translation in CJU have two main groups of modules to take: the translation related modules and the language related modules.[43] In the case of translating into the second or foreign language, this section offers a general instructional framework to incorporate language learning activities into translation practices, so that translating becomes not only an end but also a means for improving the student's ability in the second language. In doing translation exercises specifically designed for improving certain aspects of language the student not only learns general translating skills but also acquires the language ingredients intended.

As previously argued, idiomaticity is an important feature which is generally lacking in the writings of the translator as a second language learner. Since both grammar and vocabulary instruction have had long established traditions, it is imperative that we do something to help translation trainees to appreciate the idiomatic aspects of the second language. The research and future development of Translation Micro Worlds seems a viable way to help the translation learner to write idiomatically in the second language.

8.3 Translation methods tutor

The Translation Methods TMW (TM-TMW) is a more complicated system than the Jokes TMW. Although the design of this TMW still relies on the compilation and pre-processing of corpora, it appeals to more sophisticated techniques in computational linguistics and machine translation. Its instructional domain also

[43] Details of the course modules will be discussed in Chapter 11.

involves more sophisticated concepts in linguistics and translation on top of the notions of collocation and idiomaticity.

8.3.1 Introduction

It is arguable that in teaching translation, the concept of "degree of freedom" (Zhou, 1996) is one of the first important concepts the student needs to know when learning to translate. According to Zhou, there is a continuum of translation methods available between the two extremes of word-for-word translation and adaptation, which differ in the degree of freedom allowed while translating. In Zhou's model, word-for-word translation has the lowest degree of freedom because the translator is allowed the least room to manipulate structure when translating. The adaptation method, on the other hand, has the highest degree of freedom since the translator can change structure and adapt meaning at will. Zhou offers a model of translation methods on such continuum as in (3), ordered by the degree of freedom they are allowed.

(3)

- Word-for-word translation
- Literal translation
- Semantic translation
- Communicative translation
- Free translation
- Adaptation.

A similar model is also seen in Newmark (1988), who discusses the varieties in (3) under the rubric of 'translation methods'. Newmark organises the translation methods according to whether the emphasis of the translator is on the source language (SL) or the target language (TL), as shown in the two separate hierarchical lists in (4):

(4)

SL emphasis

- Word-for-word translation
- Literal translation
- Faithful translation
- Semantic translation

TL emphasis

- Adaptation

- Free translation
- Idiomatic translation
- Communicative translation

This model shows that word-for-word translation, for example, is the closest in form to the original structure of the source text; whereas adaptation puts the most emphasis on the fluency of the target text. The significance of this distinction to a translation learner should be clear – in the first place, a translator must know when to adopt which method in order for her work to fit in all kinds of demands for translation.

This section explains the development of a Translation Micro World which uses machine translation techniques and bilingually aligned corpora to help trainee translators learn how to vary their styles of translation along the "degree of freedom" continuum, i.e. how to appreciate the fine distinctions between the translation methods and to produce the target text at a desired level. But first we must explain how the translation methods are defined in the translation literature and adapted to our TMW. Note that in practice it is neither easy nor useful to distinguish so many kinds of translation methods. Newmark (1988) himself, for example, only contrasts semantic translation and communicative translation in his translation examples. In our TMW environment, we consider it appropriate to introduce four levels of translation methods – word-for-word, literal, semantic, and communicative method, as the free translation method and beyond are too difficult to grasp with a computer.

8.3.2 Translation methods

The following discusses how the translation methods are defined in Zhou's and Newmark's model, and how they are adapted to fit in the computational design of the TMW. Contrary to the previous models, the direction of translation in our discussion of translation methods is translating from the foreign language (English) into the mother tongue (Chinese). As previously indicated, our targeted users are the students in the translation department of Chang Jung University, whose ability to produce idiomatic L1 to L2 translation is highly questionable, and at the initial stage there is no possibility to teach them to vary their translations based on different translation methods in the case of translating into the second language. Hence it is considered more appropriate to introduce the concept of translation methods following the L2 to L1 direction in the first instance. After the students have a firm grasp of the main ideas behind the scheme of translation methods, they can then be taught to apply the same concepts to translation into the second language.

8.3.2.1 Word-for-word translation

For the word-for-word translation method, Newmark says, "This is often demonstrated as interlinear translation, with the TL immediately below the SL words." (p.45). Zhou, on the other hand, proposes that the word-for-word translation is to "translate each word based on the first definition of such word in a bilingual dictionary, keeping the original word order". The definition for this method is straightforward and so is the machinery for producing the translation. We only need an electronic bilingual dictionary (in this case an English-Chinese dictionary) and the translation procedure consists only of replacing each word in the source text with the first definition of such word in the dictionary. Not surprisingly, some anomalous 'translation' is inevitable with such elementary techniques. Take the Chinese example sentence in (5), which is translated into English by the word-for-word method in (6)

(5) A Chinese sentence

經濟	不	景氣,	失業	率	不斷	攀升
jingji	*bu*	*jingqi*	*shiye*	*lv*	*buduan*	*pansheng*
ECONOMY	NOT	PROSPEROUS	UNEMPLOYMENT	RATE	INCESSANTLY	CLIMB UP

(6) Word-for -word translation

Economy not prosperous. Unemployment rate incessantly climbs up.

The translation happens to be quite acceptable because the sentence structure is quite 'flat' and Chinese and English do share the same basic word order of SVO.

8.3.2.2 Literal translation

For literal translation Newmark (1988) proposes: "The SL grammatical constructions are converted to their nearest TL equivalents but the lexical words are again translated singly, out of context" (p.46). The point is then for the structure in the target language (TL) to be grammatical, eliminating the possible random, gibberish word strings produced by the word-for-word method. Zhou (1996) gives further hints for the degree of grammaticality the literal translation method should accomplish. According to Zhou, what the literal method does on top of the word-for-word method is: "In response to target language grammar, make minimum adjustment of word order and addition or omission of words, still disregarding co-text of discourse altogether." (p.24). We can then define the literal translation method

computationally as "rearranging the word order of SL to that of TL, adding or deleting functional words where appropriate to meet the minimum local grammatical requirement". This can be illustrated by the following example, where the English sentence *I read books* is translated into Japanese through two different translation methods:

(7)
(a) The source text
I read books
(b) Word-for-word translation
* *watakusi* *yomu* *hon*
I READ BOOK
(c) Literal translation
watakusi *wa* *hon* *wo* *yomu*
I 'Topic marker' BOOK 'Object marker' READ

In (7b), the Japanese translation is ungrammatical because of the incorrect word order and the lack of function words. In (7c), however, the Japanese sentence is correct as the literal translation method has rearranged the word order and added in functional words as appropriate. Let's take another example -- the Chinese sentence in (5). We now translate it on the basis of literal translation method, into the English text in (8).

(8) Literal translation
The economy is not prosperous. The unemployment rate climbs up incessantly.

Note that the function word *the*, previously not existing in word-for-word translation, has now been added, and the word order is slightly adjusted. The literal translation method is a good way of producing quick and, in many cases, minimally understandable translations.

8.3.2.3 Semantic translation

For practical reasons it seems advisable to conflate Newmark's definition for faithful translation and that for semantic translation into one single category and keep the name of semantic translation. According to Newmark, "A faithful translation attempts to reproduce the precise contextual meaning of the original within the constraints of the TL grammatical structures" (p.46). According to him, the method

" 'transfers' cultural words and preserves the degree of grammatical and lexical 'abnormality' ... in the translation". That is, it "attempts to be completely faithful to the intentions and the text-realisation of the SL writer" (ibid). More concretely put, the method tries to be faithful to both the meaning and the form of the source text, constantly seeking a good balance between the two. On the other hand,

> Semantic translation differs from 'faithful translation' only in as far as it must take more account of the aesthetic value (that is, the beautiful and natural sound) of the SL text, compromising on 'meaning' where appropriate so that no assonance, word-play or repetition jars in the finished version. (ibid.).

It seems then that Newmark's faithful translation and semantic translation methods differ not in principle but in some minor linguistic features. On the other hand, Zhou (1996) says that the translator following the semantic translation approach will "try to keep the form of the original text and maintain the intention of the author, while at the same time attempt to make the translation natural and fluent " (p. 24), which seems a good summary for Newmark's semantic and faithful translations put together.

In any case, a translation rendered by the semantic translation methods should be reasonably correct in meaning and form, with sound sentential structure and with no phrase-internal ungrammaticality. We go back to (5) again and translate it this time with the semantic translation method, with the following result:

(9) Semantic translation
The economy declines, and the unemployment rate climbs up incessantly.

A translator following the semantic translation approach should always produce a more 'fluent' translation than one following the literal translation approach, it being more concerned about reproducing the meaning with readable texts rather than clinging to the author's original form.

8.3.2.4 Communicative translation

The difference between the semantic translation and the communicative translation is: that in the semantic translation method the translator should maintain a certain degree of meaning-and-form correspondence in the source language while endeavouring to transfer the correct meaning in the source text to the target text. In the case of communicative translation, however, the translator is mainly concerned with searching for expressions in the target language to better communicate that meaning to the reader, ignoring the concerns about finding lexical or structural equivalents

altogether. In the latter case, 'readability' in the target language is a stronger concern than 'faithfulness' to the form and structure of the source text.

The computational definition for communicative translation can be derived from the definition of semantic translation, if it already exists. If the computational definition for semantic translation is something like "a translation produced by taking both semantic representation and lexical/grammatical information extracted from the source text into consideration when generating the target text", then for the communicative translation, as readability is a more important concern than faithfulness to the original structure, we would like to generate the target text based solely on the meaning of the source text, disregarding the original lexical/grammatical information altogether. Hence the computational definition for communicative translation is simply the definition for semantic translation stripped off the lexical/structural information from the SL. That is, the generator on the side of the target language relies entirely on TL rules in producing the target text.

In Newmark (1988) there is a level called 'idiomatic translation' beyond the communicative translation method. Newmark says, "Idiomatic translation reproduces the 'message' of the original but tends to distort nuances of meaning by preferring colloquialisms and idioms where these do not exist in the original." (p.47). We find it helpful in practice to incorporate the essence of idiomatic translation into the communicative translation, i.e. the 'colloquialisms and idioms' bits. We propose that the idiomatic translation be defined computationally as the communicative translation plus the refinement of collocations, idioms and 'lexicalised sentence stems'[44] in order to make the translation more fluent and native-like. Again, the Chinese sentence in (5) is translated into English, this time via the communicative translation approach, in (10):

(10) Communicative translation
The unemployment rate continues to rise with the economic downfall.

A noticeable fact about the communicative translation, as defined in our framework and illustrated in (10), should be the abundance of collocations (e.g. *economic downfall*) and the use of sentence stems (e.g. *NP continue-TENSE to VP with NP*), both of which contribute greatly to the native-like fluency.

[44] Originally proposed by Pawley, & Syder (1983). See Shei & Helen (2001) for a discussion of its implication for language learning and translation.

8.3.2.5 Free translation

According to Newmark (1988), "Free translation reproduces the matter without the manner, or the content without the form of the original.... Usually it is a paraphrase much longer than the original,... and not translation at all." (p.47). Since 'matter' and 'content' are too vague and difficult to define and to be 'understood' by a computer, and the method is 'not translation at all' according to Newmark, we will not deal with it any further in our research. But an illustration of what it could be like is supplied in (11), again translating from (5):

(11) Free translation
The current economic situation in Taiwan is worse than the time of the previous government, and the unemployment rate reaches a new high since the new government took office.

8.3.3 Text preparation

The prototype TMW for tutoring translation methods has an inherent experimental corpus consisting of the first chapter of the English novel *Pride and Prejudice* (Austen, 1813) and four different Chinese translations based on the four translation methods previously discussed: word-for-word, literal, semantic, and communicative translations. The system draws on this translation corpus for its teaching and practice materials. Before describing how it works with the corpus, we explain how the translations are prepared for use in the system.

8.3.3.1 Fully automatic MT

The word-for-word translation version is produced by fully automatic machine translation. The technique involved in doing so is quite elementary, the essential tools being an English-Chinese dictionary and a trivial computer program for searching and extracting. In (12), a sentence fragment from Austen (1813) is translated into Chinese by the word-for-word method:

(12)

a	*single*	*man*	*in*	*possession*	*of*
*一	單一的	成年男子	在...裡	擁有	...的
yi	danyi-de	chengnian-nanzi	zai...li	yongyou	de
ONE	SINGLE	ADULT MALE	INSIDE	POSSESS	'possessive marker'

a	*good*	*fortune*
一	好的	財產
yi	haode	caichan
ONE	GOOD	FORTUNE

The result of word-for-word translation in the case of (12) is an ungrammatical Chinese phrase because, for one thing, the phrase *in possession of* is broken down to its ingredients and translated separately, and of course the result will not be acceptable in any target language. Also in Chinese a specific 'classifier' must precede the noun when a number word is used. So for English *a fortune* we must say *yi-bi caichan* instead of **yi caichan*, where *bi* is the correct classifier for the noun *caichan*.

Another translation method which could be done by automatic MT is the literal translation. Previously we have defined the literal translation method as "rearranging the word order of SL to that of TL, adding or deleting functional words where appropriate to meet the minimum local grammatical requirement". In preparing literal translations for the corpus in the TMW, the basic requirement is a series of word-order rules, which work on part-of-speech tagged Chinese sentences. The system runs the result of the word-for-word translation through these rules and correct any assumed errors in word order. On top of the word-order rules, a series of simple and local rules are proposed to handle these "minimum adjustments". For example, there is a classifier-adding rule which, when encountering a quantity word like *a*, will look up the noun following it, and retrieve (from a precompiled table) the correct Chinese noun classifier for the noun. For the example in (12), we would then have the more appropriate *yi-ge nanzi* rather than the unusual *yi nanzi* if the classifier *ge* is added, and *yi-bi caichan* rather than **yi caichan*. Another example of these local rules are the collocational rules, which examine certain word combinations to see whether they are habitual co-occurrences in Chinese. For example, a collocational rule would locate the word before *de* (which is likely an adjective) along with the word after the same *de* (which is likely a noun) and search a collocation database to see if they partially match an adjective-noun collocation in Chinese (see e.g. Shei & Pain 2000 for the processing of collocations). For the example in (12), *danyi* and *chengnian-nanzi* are extracted from the structure *danyi-de chengnian-nanzi*

and checked against the adjective-noun pairs in the collocation library. As a result, the correct collocation *danshen nanzi* ('single man', or 'bachelor') is found, which happens to be the correction translation as well in terms of meaning. The net result of the literal translation for the English text in (12) is thus:

(13)
(a) Source text
a single man in possession of a good fortune

(b) Literal translation

?一個	單身	男子	在	一筆	好的	財產的
yige	danshen	nanzi	zai	yibi	haode	caichan-de
A	SINGLE	MAN	AT	A	GOOD	FORTUNE'S

擁有	裡
yongyou	li
POSSESSION	INSIDE

The Chinese sentence in (13), having gone through the word order adjustment and some 'local improvements', reads much better than that produced by the word-for-word translation in (12). However, its overall grammaticality is still questionable, and the sentence should read remarkably similar to that produced by a novice English-to-Chinese translator, i.e. one that bears "strong marks of translation."

8.3.3.2 MT with post-editing

The semantic translations are produced via the MT with post-editing methods in this system. According to Somers (1988): "Post-editing consists of tidying up the raw output, correcting mistakes, revising entire, or, in the worst case, retranslating entire sections." (p.138). With current MT technology dealing with Chinese and English, it is still not possible to get fully satisfactory translations automatically and heavy post-editing is usually required to render semantically and formally correct translations.[45] Nevertheless, for saving human resources and for the sake of consistency when constructing our TMW, efforts should still be made to improve the performance of the MT system and to generate translations as close to the 'real thing' as possible.

[45] See Hutchins (1999) for comments on current state of MT. Also visit http://www.netat.net to get the feel of the performance of current English-Chinese MT technology.

The characteristics for semantic translation are: accurately expressing the meaning of the author and not deviating too far away from the original words and structure used by the author. Owing to the limitations of the current technology in semantic analysis, a reasonable way to emulate a true semantic translation is first doing a superficial parsing and then using an example based approach to get translation equivalents for each constituent analysed. For example, the parse of the sentence fragment in (12) could be something like: `[a single man] [in possession of] [a good fortune].` We can then deal with the translation of this sentence fragment phrase by phrase, with better hopes that the entire text is translated into idiomatic bits in the target language. Some of these phrases could be looked up in a dictionary. For example, looking up a representative Chinese-English electronic dictionary with *in possession of* yields the meaning *yong-you* ('possess'), which is much more satisfactory than extracting the meaning of its components individually as was done in the word-for-word translation. As for the phrases which could not be found on the dictionary, one way to translate them is to look up a large bilingually aligned corpus and see if we could find them there. If the dictionary checked is comprehensive and the bilingual corpus large enough, the semantic translation version of (12) should look something like (14b)

(14)

(a) Source text

a single man in possession of a good fortune

(b) Semantic translation

一個	單身漢	擁有	一大筆	財富
yige	danshenhan	yongyou	yidabi	caifu
A	BACHELOR	OWN	A LARGE SUM OF	WEALTH

There will be phrases which can not be found on either a dictionary or a corpus, which need to be filled in by the system administrator (i.e. the instructor preparing the corpus for use in the TMW), using the same criteria for semantic translation. Post-editing is also required to correct any syntactic or lexical anomalies produced by the machine via the above method.

8.3.3.3 Human translation

Currently the easiest, if not the only possible, way to find communicative translation for a given text is to look for works produced by professional human translators. For

example, one existing Chinese translation of *Pride and Prejudice* produced by established translators can be found at Ocrat Chinese Pages (n.d.) and the translation it offers for the sentence fragment in (9) is simply:

(15)
(a) Source text
a single man in possession of a good fortune

(b) Communicative translation
有錢的　　單身漢
youqiande　danshenhan
WEALTHY　BACHELOR

In fact, this is the only version so far which correctly identifies the source structure as an NP (noun phrase) – all previous translations look like some sorts of Chinese clause. Also it correctly does away with the article *a* in the source text, which is often unnecessary for an NP in the Chinese discourse. If the instructor wishes to use a corpus for which there is no existing human translation for the source text, then the instructor would have to translate the corpus herself in a professional way, i.e. the translation must be "comprehensible and pleasurable to read" (in Zhou's words) to qualify as communicative translation, on top of being correct in meaning.

8.3.4 Applications

Now that the corpus is properly organised, we go on to explain how the system works to put the source and translation texts to use. The basic form of application is for the system to be used to illustrate the existence and consequences of degrees of freedom in translation. Further application sees the basic machinery incorporated into more complicated systems for teaching certain translation skills.

8.3.4.1 Basic form

The components and work flow of the Translation Methods TMW are as follows:

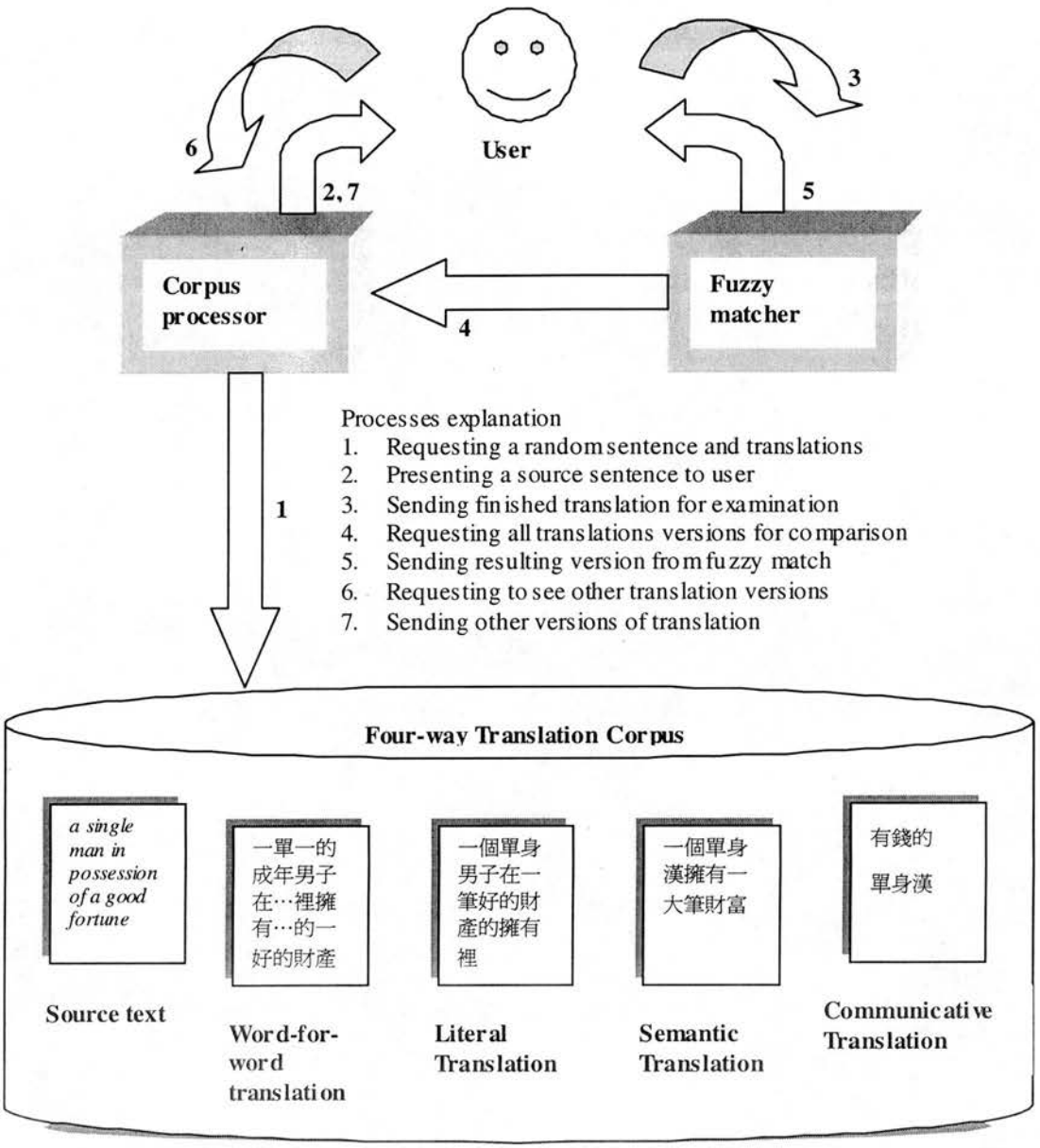


Figure 8-5: The Translation Methods TMW -- modules and work flow

In the basic working mode, the system offers an English sentence for the student to translate. Upon finishing, the student submits the translation and the system compares it with its pre-stored versions of translation for the same sentence, decides on the closest match based on a matching algorithm like that used by translation memory software (See 7.2.4 for explanations). That is, the system calculates the similarity measure between the student translation and each of the pre-stored translations for the same source text based on a word-matching algorithm, and decides which of the pre-stored translations is the closest to the student translation. The best match will be presentsed to the student together with the translation method tagged with that version. The student if she so desires can click on any other version of the

translation to compare and see if the system has made a fair judgment about the category of the student's translation, as well as to appreciate the differences between the results of different translation methods. Figure 8-6 is an interface where all the above incidents happen.

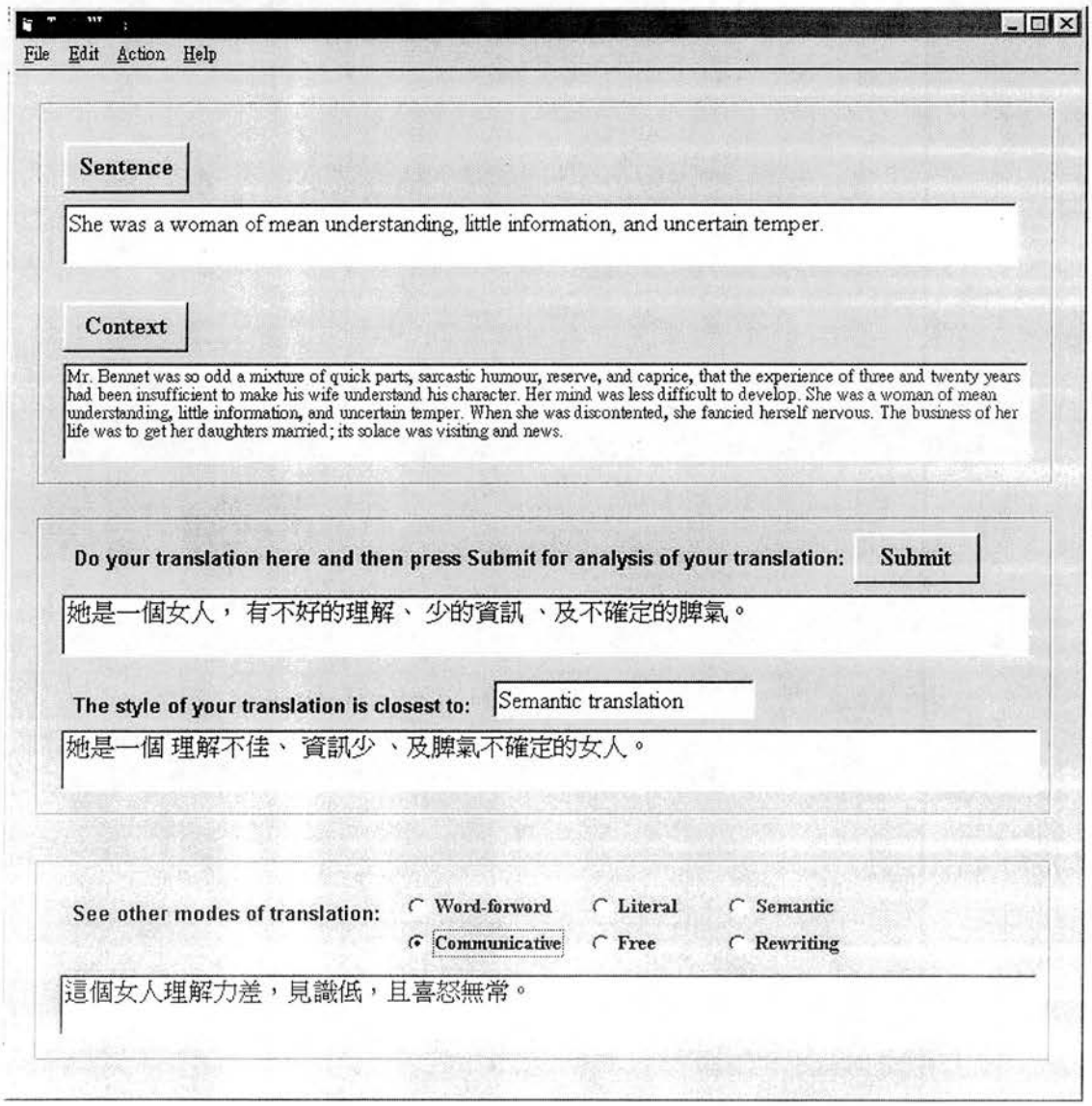


Figure 8-6: An example interface of TM-TMW

In this very simple interface, the translation students press the SENTENCE button to get a source sentence for translation. A further CONTEXT button could be pressed to access more surrounding sentences in order to get contextual information. After the translation is done and submitted to the system, the system judges the kind of translation method being used and retrieves the pre-stored version of translation under the same translation method. Finally, the student can call out any version of translation of the same source sentence and observe their differences. Note that

although free translation and rewriting are not included in our computational model, the system does not preclude the possibility that the instructor may wish to include them, and offers a chance for the system administrator to do these extra versions of translation manually and include them in the pre-stored translation corpus.

8.3.4.2 Integrated environment

The second stage of research allows the basic application to be integrated into another computer system for teaching translation skills. For example, one of the popular methods for teaching translation is through contrastive analysis (CA; see Hoey & Houghton, 1998 for a discussion on CA and translation). The curriculum following this approach usually constructs the syllabus units based on parts of speech (Wu, 1982). In each instruction unit a part of speech (e.g. conjunction) is analysed contrastively in both languages in order for students to understand the behavioural change of this part of speech when the translation task involves language bits belonging to this category. When focusing on the translation of a certain part of speech, it is also interesting to see how its behaviour can change following the change of translation methods. For illustration we take an example from Wu (1982):

(16)
(a) Source text
We are not gentle folks and troubadours

(b) 'Bad' translation

我們	不是	紳士	和	行吟	詩人
women	bushi	shenshi	han	xingyin	shiren
WE	NOT-BE	GENTLEMEN	AND	STROLL-SING	POET

(c) 'Good' translation

我們	不是	紳士	也	不是	行吟	詩人
women	bushi	shenshi	ye	bushi	xingyin	shiren
WE	NOT-BE	GENTLEMEN	ALSO	NOT-BE	STROLL-SING	POET

Wu considers (16b) a bad translation as the direct translation of *and* causes an unusual usage of *han* in Chinese. On the other hand, (16c) is a better translation because a repetition of *bushi* is used for the double negation, which is a more common usage in Chinese. To use the terminology of translation methods, we can explain that (16b) is bad because the translator uses the word-for-word or the literal translation method,

which may not be a good strategy for dealing with conjunctions in translation; (16c) is good because it reflects the spirits of semantic translation method, probing into the meaning of the source text while maintaining a minimum structural correspondence (*bushi...ye bushi...* is not a 'translation equivalent' of English *and*, but is a kind of conjunction, the correlative conjunction, in Chinese). The basic system incorporated into an TMW for CA-based translation teaching can have the interface illustrated in Figure 8-7.

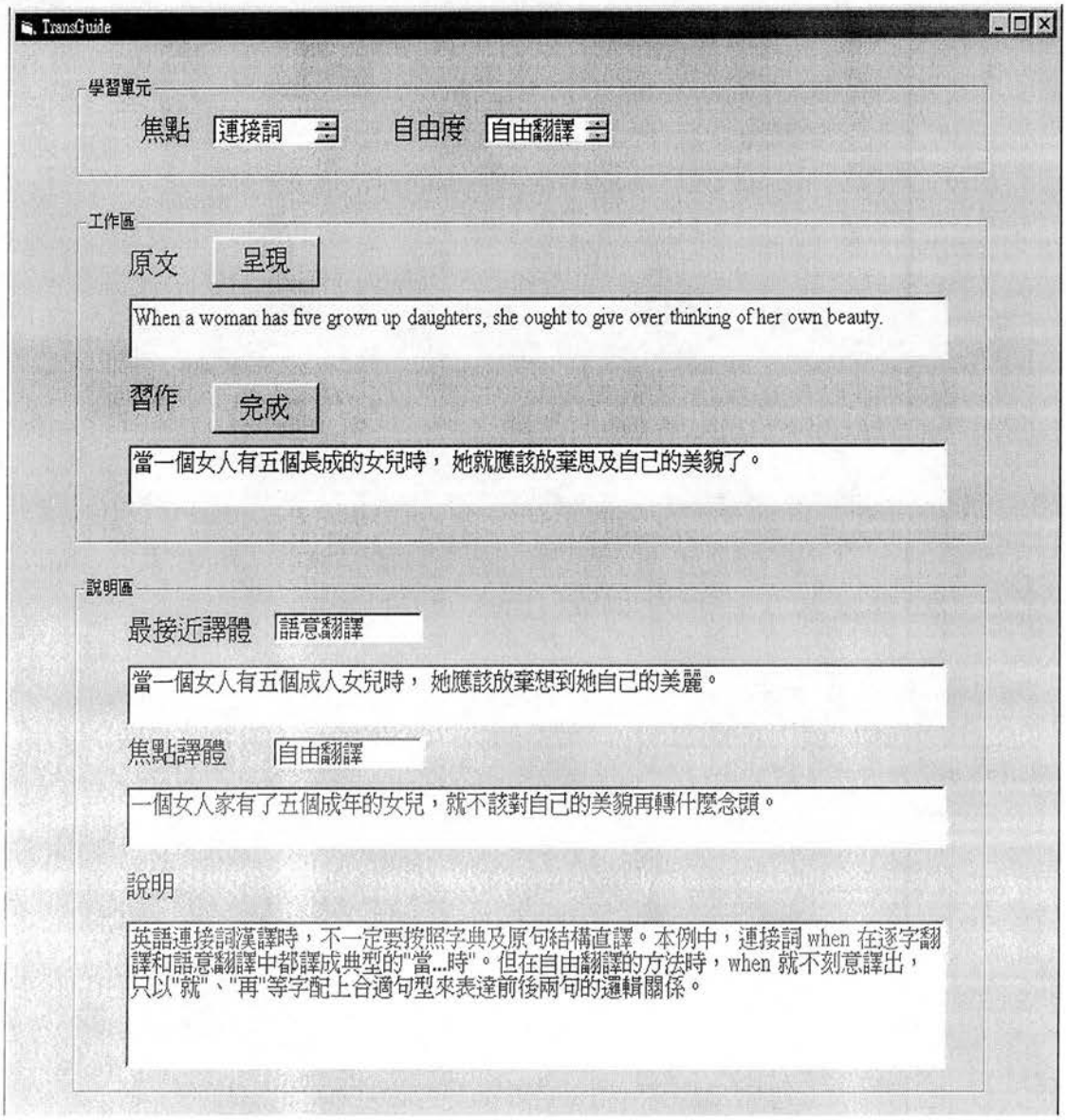


Figure 8-7: The integration of TM-TMW into another system

The system lets the student choose which part-of-speech or structure he wants to work on. A sentence in the source language containing the desired word or structure is offered for translation. The student translates the sentence and then presses the

"Finish" button to see three things in the "Explanation" area:

1. the translation of the source sentence closest to the student's translation offered by the system,
2. the translation of the source sentence in a particular category of degree of freedom selected by the student, and
3. the explanation of how that particular part-of-speech or grammatical structure is dealt with generally and in terms of different approaches to degree of freedom which are canned text supplied by the instructor and pre-stored in the system.

The role TM-TMW plays in this kind of system is to add in another dimension for consideration, allowing the concept of degree of freedom or translation methods to be integrated in the teaching of various skills in translation. For example, Liu (1997, p.218) suggests five ways for translating English prepositions into Chinese: translating into verbs, translating into prepositions, translating into negatives, translating into modifiers, and translating into idioms. TM-TMW can relate these techniques to the principles of translation methods and enhance the teaching activities in terms of their richness and clarity. For example, in translating the phrase *to work against time* into Chinese, we can say not only that there are five methods for tackling the preposition *against*, we can also say that using the first method corresponds to the communicative translation approach, the third method to the literal approach, and so on and so forth.

8.3.5 Remarks

This section considers a TMW which acts as an illustrator of the working of a fundamental concept in learning to translate – the principles behind the variation of translation methods or the criteria by which the translator decides on the balance between being faithful to the source text and creating a readable target text. It also explains the possibility of the TMW being integrated into a larger system for teaching translation skills. For a novice translator to have this kind of recognition is important in the beginning and throughout her translation career, as a lot of translation techniques can be explained, at a certain level, by the principle of translation methods.

8.4 Summary and conclusion

This chapter introduces two pieces of Type B software, or two Translation Micro Worlds, both of which make extensive use of bilingual corpora and materials generalised from corpora. In the case of the Jokes TMW, collocations and lexical phrases are extracted from the jokes corpus and used as reference materials for

in the post-communicative world' and who observe that "there has, recently, been a shift in emphasis back towards the teaching of structure", the 1990s saw a 'grammar boom' in language teaching and second language acquisition (SLA) research. The latest development includes grammar being integrated with multimedia features to add to the fun and effect of learning (e.g. *Focus on Grammar* reviewed by Jung, 2000), with the Internet to provide fresh online learning experiences (e.g. The Internet Grammar of English),[46] and with corpora to offer the learner tremendous amounts of authentic materials for generalisation and emulation (e.g. McEnery & Wilson, 1997). With this revived interest in grammar, one wonders whether translation is going to regain a place in language teaching also, in the years to come. Just as one argues that CLT fosters fluent but inaccurate language learners, and therefore explicit instruction of grammar is required, so one can argue that translation is an activity which forces the learner to comprehend or produce the language accurately, and therefore translation exercises can help the learner improve her grammar and hence the target language as a whole. That translating demands accuracy is because the learner, in the face of a stretch of source language text to be translated, needs to consider all the candidate words and structures coming to her mind and clarify the meanings and appropriateness of such by referring to all sorts of tools, and finally settle upon a stretch of target language text which she thinks best represents the original text. No other activities can seem to achieve effects on language processing equivalent to those demanded by translation. For example, in spontaneous writing or speaking activities, the learner can always choose the words and structures she is confident in using and avoid any uncertain ones. But in translation the fixed source text demands that the learner consider unfamiliar structures and thereby forces the learner to find out and learn the usage of new words or structures.

9.2 Integrating translation

The discussions in this section centre around how translation can be incorporated into existing CALL models. Based on the recent trends of development in CALL, four categories of CALL software are distinguished here, each of which is separately discussed according to its compatibility with translation. Note the four categories of CALL distinguished here -- intelligent, multimedia, corpus-based, and Web-based CALL -- are not mutually exclusive. That is, a piece of CALL software can be multimedia and Web-based, for example, or it can be an intelligent, corpus-based system. Also, the classification does not intend to exhaustively cover all CALL possibilities, but only dwells on the more recent and interesting varieties.

---

[46] http://www.ucl.ac.uk/internet-grammar/

9.2.1 Intelligent CALL

Intelligent CALL (ICALL) is a kind of intelligent tutoring system (ITS) which concentrates on the teaching of language. According to Burns & Capps (1988), an ITS is an integration of the following modules or elements:

- **The expert module**: This is the ITS data structure where the expert knowledge in the domain of teaching is encoded.
- **The student diagnosis module**: This is a knowledge structure which is able to decide the student's current state of learning.
- **The curriculum and instruction module**: This is responsible for selecting and sequencing the teaching material, answering students' questions, and deciding on when students need help.
- **The instructional environment**: This refers to the activities, situations and tools offered to the student as support for using the system.
- **The human-computer interface**: This could include features like natural language interfaces, speech recognition, touch screen technology, three-dimensional graphics, etc.

An area where ITS is enthusiastically embraced in language learning and teaching is the automatic recognition and correction of students' errors in language production. As can be seen from Bolt's (1993) analysis of several examples of such programs, the error detection technologies are still far from being mature. However, if we import the principles of the Translation Micro World into the domain of error recognition and correction, there may be some improved results. The idea is to let the students do translation exercises with fixed materials and record all the errors that ever occur with each user of the system. Errors will be individually recorded and generalisations will be made from the individual errors. As recorded errors accumulate from different users of the system, more and more errors can be reliably predicted, which adds to the width of the system's error processing power. When a corpus in the TMW is exhaustively exploited (i.e. all potential students have used the system), a new corpus can be introduced for further practice and to collect more error tokens and types. As more errors of similar types occur, generalisations can be reliably made, which adds to the depth of the processing power of the system.

Some ITSs of the error detection type already make use of translation in their instruction processes. Earlier we mentioned the construct of Wang & Garigliano (1992), which uses a dual set of English and Chinese grammar to detect learners'

errors in doing Chinese-to-English translation exercises for students learning Chinese as a foreign language. Wang & Garigliano's is a grammatical rule-based system which actually has a quite limited scope (going only as far as the set of machine operable grammatical rules goes) and suffers the disadvantages of all such systems, i.e. erroneous judgement made by the machine about the user's input based on inadequate or inapplicable rules, which can cause learners' confusion or even misunderstanding about the target language. The use of an *error library* as mentioned above, on the other hand, offers a likely solution to more intelligent diagnoses of learners' input. Like Wang & Garigliano, translation is used to elicit students' language productions and an attempt is made to track down errors caused by language transfer in the learners' interlanguage, but unlike Wang & Garigliano, the TMW approach uses a context-rich corpus for the translation activities, rather than individually constructed sentences, and bases the diagnosis of errors on actual data collected and rules generalised from these errors.

9.2.2 Multimedia CALL

A multimedia program is a rich combination of video, audio, graphics and text and is arguably the most commonly adopted form of commercially produced educational software. Fleta et al. (1999) conducted a questionnaire survey of the application of four kinds of commercial multimedia software for learning English in a university course, *Interactive Computer Practice of English*, and found that students evaluated the multimedia programs highly in the areas of listening comprehension and vocabulary and grammar learning. The students also confirmed their interest in the content of the multimedia programs and the improvement they made throughout the course, two factors which contribute greatly to students' motivation for learning. Harben (1999), on the other hand, found multimedia-based CALL design to have two pedagogical advantages for listening comprehension exercises: it provides the learner with authentic listening experiences by inducing the learner to do both bottom-up and top-down processing, and it offers flexibility in self-learning and thereby fosters learner autonomy.

Chapelle (1998) proposes seven criteria for developing multimedia CALL based on second language acquisition theories:

1. The linguistic characteristics of target language input need to be made salient.
2. Learners should receive help in comprehending semantic and syntactic aspects of linguistic input.

3. Learners need to have opportunities to produce target language output.
4. Learners need to notice errors in their own output.
5. Learners need to correct their linguistic output.
6. Learners need to engage in target language interaction whose structure can be modified for negotiation of meaning.
7. Learners should engage in L2 tasks designed to maximize opportunities for good interaction.

Excepting 6 and 7, which dwell on simultaneous interactivity that is inherently contradictory to the activity of translating, we can expect translation to contribute to the other five designing principles of multimedia CALL. Namely, translating from L2 to L1 can satisfy the criteria of 1 and 2; whereas translating from L1 to L2 can satisfy 3, 4, and 5. It is arguable, as a matter of fact, that translating may be an even better way for students to notice errors (4) and correct their output (5) than plain writing, if this is conducted in a TMW environment as described in Chapter 8, where a model text in the target language can always be called out for comparing and emulating.

For example, Brett (1997) introduces DIY multimedia software using the "worksheet" approach. The worksheet is a (now ancient) Microsoft Word 2 document consisting of text, sound, pictures and video clips played by the Windows Media player. One example Brett gives is a reading-vocabulary task where students work in pairs to read an article and then do multiple-choice questions on reading comprehension and vocabulary; do follow-up discussions, watch a video clip, and discuss again. Brett correctly singles out the importance of 'language awareness' and suggests "tasks that followed up skills work by focusing on some grammatical areas exemplified by the input". However, if we contrast this kind of design with Chapelle's (1998) criteria for designing multimedia, we find that the elements of language output and its processing (criteria 3-5) are severely lacking in the worksheet approach. This is arguably a good opportunity for translation exercises to step in, which have the potential to satisfy Chapelle's 3-5 criteria by providing a chance for the learner to produce L2 output, and offering feedback whereby the learner can identify the likely errors in her output and make corrections.

9.2.3 Corpus-based CALL

The corpus-based approach is essentially a more advanced mode of instruction and is best suited for intermediate to advanced students. This point can be seen from McEnery & Wilson's (1997) characterisation of corpus-based learning by reference to

four features:

- **Discovery learning**: The learner navigates through the corpus under her own impetus with a certain task in mind and finds instances of language in the corpus which verify or overthrow her hypothesis.
- **Divergent learning**: This means that "different students take different paths through the data and find slightly different things", in McEnery & Wilson's original words.
- **Mediated learning**: The student does not learn directly from the content presented for learning (as in a textbook) but instead learns indirectly from analysing the language in the corpus.
- **Directed learning**: The teacher plays a guiding role rather than a traditional instructing role and "it is the students who lead themselves and one another through the learning process".

By far the most frequently seen corpus-based CALL software centres around the use of a concordancer (Johns, 1997; Gavioli, 1997; Whistle, 1999; and Cobb, 1999). Cobb (1999) argues that a concordancer-based program transcends traditional learning methods for lexical acquisition in that it can offer both breadth ("explicit learning of words on lists") and depth ("implicit learning of words through extensive reading") in learning a language. In other words, while the traditional approaches can only concentrate on either breadth or depth at a time, the concordancer backed up by contemporary computer technology can deal with the breadth-depth paradox competently. That is, a concordancer offers students opportunities both to customise their own learning list while navigating through the corpus (breadth) and to go into deeper reading of the corpus by expanding the concordance lines (depth).

A concordancer-based CALL framework is a good host for translation exercises, which in return can add to the program's usefulness in ways as previously described, i.e. forcing the learner to use unfamiliar words and structures and to be accurate in using or comprehending language, etc. As demonstrated in Chapter 8, bilingual corpora are infinitely useful for teaching translation. For a concordancer, novel ideas like 'bilingual concordancing' can also be explored to expand the learning possibilities. That is, when the learner requests the concordancing of a certain word in the target language, the system has the ability to return not only the concordance lines of the word in L2 (e.g. English), but also concordance lines in L1 (e.g. Chinese) based on the translation equivalent of the input word.

We could not look at photographs or **read** her old letters. Every knock at the
photo in the magazine, I felt I could **read** her thoughts. Thoughts that were
all blood runs down their chests. I've **read** how the fighting gloves get heavy
higher order of Catholicity must be **read** in the context of the peculiar It
was therefore extremely galling to **read** in LPN about Claire Rayner's
of thing). Nor, probably, would I ever **read** it again and even the above
His parents taught him to **read** music when he was only four, and
favoured potential jurors who did not **read** newspapers or watch television.
going on at this time: It is sad to **read** of the selfishness of many people
suggestion. Witches' confessions were **read** out at their executions, public

忽然說道："你妹妹現在還在那里 **讀**書麼？"曼璐只含糊地哼了一聲，懶得回答
後回鄉下去了，慕瑾仍舊留在上海 **讀**書，住在宿舍里，曼璐和他一直通著信，
鵬，一鵬從前和世鈞一同到上海去 **讀**大學的，因此和叔惠也是同學，但是因爲
子看待，並不怎麼注意。他在上海 **讀**書的時候，專門追求皇後校花，像翠芝這
都是同學，他說叔惠那時候是一面 **讀**書一面教書，因爲家里窮。石太太當時聽
，有翻譯的小說，也有她在學校里 **讀**的教科書，書脊脫落了的英文讀本。世鈞
很別扭。別的還在其次，第一就是 **讀**信和寫信的環境太壞了。曼楨的來信寄到

For example, from the English concordance lines above, the student will learn that *read* takes *letters*, *thoughts*, *music*, *newspapers*, etc. as objects in English. Whereas from the Chinese concordance lines, the learner will notice that *du*, the Chinese translation equivalent of *read*, collocates with *book*, *university* and *letter*, etc. And the learner can start wondering whether in English *read* takes the translation equivalents of these as objects too,[47] or if there are other verbs collocating with them, and so on.

A similar approach to the above is taken by Wang (2001), who discusses how *E-C Concord*, an English-Chinese parallel concordancer, can be used for Chinese students learning English. Wang uses groups of parallel concordance lines the Chinese parts of which all contain the Chinese word *xian4zai4* ("now") and the English ones, its English equivalent (i.e. *now*). Each group of parallel concordance lines illustrates a pattern where the use of Chinese *xian4zai4* is different from its English translation used in an English sentence. Thus from these pairs of bilingually aligned sentences the student observes and generalises the behavioural patterns of certain lexical items in English by noting the differences between Chinese and English with regard to the use of these items. Wang suggests that this is an effective

[47] Already we find a contrast in these few bilingual concordancer lines: In Chinese *du daxue* ('read a university' literally) is an acceptable collocation; whereas in English it is not possible to say so.

way for students to gain in-depth lexical knowledge in the target language.

9.2.4 Web-based CALL

A Web-based CALL program would exist in a Web-based classroom. According to McCormack & Jones (1998): "A Web-based classroom is an environment created on the World Wide Web in which students and educators can perform learning-related tasks" (p.1). McCormack & Jones suggest that the Web offers the following benefits to education:

- **Computer mediation**: The vast pool of information resources can be accessed with a computer.
- **Geographic independence**: Information management is no longer limited to the physical buildings of the institution and can be stored, updated and distributed freely online.
- **Temporal independence**: Asynchronous communication allows both student and educator to access the course at any convenient time and formulate questions or answers without time pressure.
- **Platform independence**: The Web programs can be designed in such a way that they can be accessed by most Web browsers regardless of types of computers or operating systems.
- **A simple, familiar, useful interface**: The popularity of the Web means most students know how to use a Web browser and related tools already, reducing the students' efforts in learning new skills for learning.
- **Increased communication**: This is especially true for Chinese students, who seem to talk a lot more electronically[48] than in class with peers or the instructor.
- **Increased learner control**: The combination of various flexibility features of the Web-based learning environment gives the learner the feeling of control over their study.

The Web-based approach often involves the coming together of practitioners across different fields. Hence a systems approach for a business English course proposed by Curtis et al. (1999) is logical, where a course development team consists of: a project manager, a language instructor and tutors, an advisor from the Faculty of Commerce, an Internet expert, a graphics designer, an AV team and a computer programmer, a market researcher, an advertising agent, and an evaluator.

[48] Mostly via BBS for Taiwanese students, which is the most popular chatting and posting program in Taiwan.

Furthermore, Web-based learning has become a trend and now often a joint adventure between academic institutions (Sheely et al., 2001).

Literature about specific advantages or disadvantages of incorporating translation in a Web-based course is still scarce. For example, although Connell (1999) introduces a distance learning translation course, the article concentrates mostly on discussing general features of distance learning, with only a translation-specific idea about the keeping of a 'translation log' which seems not necessarily connected exclusively (or favourably) to the Web-based approach. To the present author, incorporating translation exercises in a Web-based CALL structure has at least the following additional advantages on top of the general benefits of a Web-based CALL module:

- **Electronic tools and resources**: The learner can access not only the course materials, but can also visit any online Web pages to use a tool related to translation and language learning (e.g. online concordancing or machine translation) or to draw resources (e.g. online corpora, electronic books and journalism) and to acquire language skills through using these tools and resources for translation tasks.
- **Feedback management**: With a Web-based approach only, the instructor can analyse all students' translation works and post valuable generalisations (about errors in using the target language etc.) conveniently on the Web space for all course participants to see and improve their L2 (See 11.2.3.4 for further discussion).

There are some merits, then, for translation teaching to be subsumed in a Web-based CALL framework. However, as pointed out previously, a Web-based course is often team work and a joint adventure, which demands much more complicated efforts than, say, a stand-alone intelligent multimedia CALL system. To add translation means that at least one more expert (on translation) is needed in the Web-based CALL development team and possibly the inclusion of one or more translation-related institutions.

9.3 An Example

An example is offered here as to how translation can become a helpful element in the range of CALL software discussed above. This involves the integration of a translation support mechanism into a foreign language learning CALL environment, in order to reduce the learner's anxiety in working with an unfamiliar language.

9.3.1 Introduction

One of the negative factors in learning a foreign language (FL) is the learner's anxiety towards working with the FL in class. This includes both producing and receiving the foreign language. Ellis (1994) observes, "When anxiety does arise relating to the use of the L2, it seems to be restricted mainly to speaking and listening, reflecting learners' apprehension at having to communicate spontaneously in the L2" (p. 480). For most Chinese students, to communicate comfortably in a foreign language is not an easy task due to various psychological and environmental factors, especially those imposed by the traditional educational system. Thus communicative language teaching (CLT) is not a particularly successful methodology in the Chinese world (Anderson, 1993; Cortazzi & Lixian, 1996; Dirksen, 1990). In its extreme form, the reluctance to communicate in the foreign language even results in hostility in seeing any Chinese people talking in English (Lai, 1994). A likely way to overcome the anxiety or abhorrence in using a FL is to offer "translation support" to the learner during communication when needed. This practice (of resorting to first language and translation) has already been a tradition of some humanistic approaches such as the Community Language Learning (see Richards & Rodgers, 1986), which manifests its practical value in the field.

A possible framework of intelligent language tutoring is discussed here focusing specifically on the learner's affective domain with regards to the provision of L1 support during the process of FL learning. While explaining the affective domain in some detail, the framework leaves open the part of the cognitive domain for any future models to fill in, although one or two scenarios are described as an example of how the affective domain should work.

A system under this intelligent instruction framework would work this way: as the learner starts using the system, he is given a diagnostic test to decide on his cognitive profile – how good he is in the foreign language, and a diagnostic questionnaire to decide on his affective profile toward the foreign language – how he feels about the language (Bull et al., 1993, takes a similar approach). Based on these two sets of data, the system makes judgments as to the learner's overall profile and constructs a learner model, and decides on the instruction strategy, mainly on how to give first language support during the learning process. The general principle should be to move the learner from the affective domain (the use of L1) to the cognitive domain (the use of FL) by gradually stripping off the learner's reliance on L1 and building up his confidence in using the FL.

9.3.2 L1 and Translation in SLA

There is no established theory as to whether the use of L1 is helpful or harmful for second language acquisition (SLA). As Stern (1983) comments, this is a "century-old debate" (p. 402) and "neither in second language learning research ... nor in foreign language pedagogy has this issue so far been resolved" (p. 403). In TESL/TEFL (teaching English as a second or foreign language) tradition, the use of L1 and translation in class has been adopted or abandoned according to the teaching methodology in question (see Celce-Murcia, 1991 or Richards & Rodgers, 1986). Affectively, however, the use of native language in a foreign language class is one of the ways to reduce the level of anxiety and get the learner into a more receptive mode. Although using translation for language learning has long been out of fashion since the "banishment" of the Grammar-Translation method, some L2 learning methodologies continue to adopt translation as a supportive or even dominant learning activity. According to Celce-Murcia (1991), in the Affective-Humanistic approach, "translation may be used heavily in the initial stages to help students feel at ease" before it is gradually phased out.

Stern (1983) points out:

> Affectively, the second language learner has to come to terms with the frustrations of non-communication. The lack of language contact and of means of expression and the absence of a safe reference system give the learner an initial intellectual and emotional shock ... (p. 398).

The value of a learner's first language for translating into and from the second or foreign language being learned lies, in this context, in its being able to offer a starting point for communication, a safe reference point, an intersection between L1 and L2 where the universal components of language apply. Other things being equal, a FL classroom with L1 support should produce more confident learners acting at ease and more willing to explore the new language.

9.3.3 MT for Language Learners

The use of machine translation (MT) in language learning has not yet received the attention it deserves, probably because the current output of MT in unrestricted domains has not reached publishable quality yet (Hutchins, 1999) and so cannot be considered to be of help to language learners who in principle should be exposed to good, representative samples of the target language. However, learners themselves

do not seem to mind the poor quality of the MT output and still make use of whatever becomes available in some FL learning situations. Yang & Lange (1998), for example, report on the language learner happily finding answers to their FL exercises through the free MT service offered on the Web. The other example is for language learners (at least the majority of college students I see here in Taiwan) to use their portable electronic translators frequently when engaging in language learning activities, whose output for sentences seems no better than crude word-for-word translation.

As for research in the area of applying MT to L2 or FL learning, reports have been scarce. A good example of applying imperfect MT output to language teaching is Anderson (1995), who reports on a Hebrew-to-English machine translation system used for English speakers to learn Hebrew. Because the MT system produces erroneous English sentences, the learner is asked to check a reference bilingual corpus and other tools to find out why the MT system makes such mistakes. We can imagine that the more sophisticated MT technology becomes, the more possibilities there are for MT to be integrated into a language teaching curriculum.

In the current language instruction framework, MT is contrived as a translator or interpreter directly of service to the FL learner. The framework anticipates that in future the MT technology will gradually approximate human-like translation. Before that can happen, however, it will help if the MT system has a "self-retrospection" capacity, i.e. being able to tell the user how good the translation being produced is. The idea is for the learner to know how the MT output can be useful to him. There should at least be three categories of translation that such a MT system distinguishes:

- Word-for-word translation: This is a minimum requirement for the MT system working under this framework. All the words in the source text should be replaced by the translation equivalents of the target language, although no further adjustment need be made. A comprehensive bilingual dictionary is required to satisfy this criterion
- Literal translation: As one might imagine, word-for-word translation can sometimes create very bizarre "sentences", especially when the two languages in question are very different. Some grammatical adjustments, like the rearrangement of word order, are necessary, to create minimally acceptable sentences. According to Newmark (1988), one of the characteristics of literal translation is that the "SL [source language] grammatical constructions are converted to their nearest TL [target language] equivalents" (p. 46). Thus literal translation works at the level of phrase structure constituents, which

should abide by the phrase structure rules, including the highest S symbol. The minimum requirement by the literal translation criterion is then for the sentence to be grammatical. This is not an easily attainable goal for MT. However, the current framework considers any approximations from contemporary MT useful to the FL learner.

- Idiomatic translation: In the framework being proposed, this is the most acceptable form of translation which improves on literal translation in the respect of idioms, collocations, discourse structures and other pragmatic or cultural factors. This should be a sentence or utterance normally used and accepted by the native speakers of the target language. For MT this is a distant goal (see Hutchins, 1999). However, if the framework is applied to a Translation Micro World discussed before, this could be produced from a pre-aligned bilingual corpus incorporated into certain designs of intelligent tutoring systems.

In learning to write or to converse in the FL, for example, the system helps the student with translation while the student is doing online conferencing or writing a composition. The student inputs a sentence in his native tongue, and the system responds by offering a translation into the target language the best it can. While anticipating the appearance of fully automatic high quality (FAHQ) machine translation in the future, the framework can make use of an MT system at any stage of perfection since by design the system should tell the user via what translation method a given translation is rendered. That is, the system should be able to tell the learner whether the translation in question is a crude word-for-word translation, a minimally acceptable literal translation, a fluent semantic translation, or a fully idiomatic communicative translation.

The student accepts and edits the machine's translation according to the translation method specified. For example, for a word-for-word translation version, the student knows that he has to do plenty of editing to make it useable in communicating. For a communicative translation, the students can be sure that the translation is instantly useable for communication (in the case of L1-FL translation) or reliable for helping to understand the target language (for FL-L1 translation). Further, the system works like a translation memory system (TMS), remembering every pair of source sentence and MT-generated, user-edited target sentence. In future, when the user requests translation of the same or similar source sentence again, the previously accepted translation will be offered.

9.3.4 An L1-Support Framework

The framework aims to take care of the FL learners' affective domain by tackling the issues of anxiety, fear, etc. in using the foreign language. The solution offered is for the system to translate from L1 to FL on behalf of the learner in the case of language production, or to translate from FL into L1 in the case of language comprehension. The idea is for the learner to have a safe place to fall back on (i.e. to the use of native tongue) while working with the foreign language, as he gradually and painlessly overcomes the language barrier between L1 and FL.

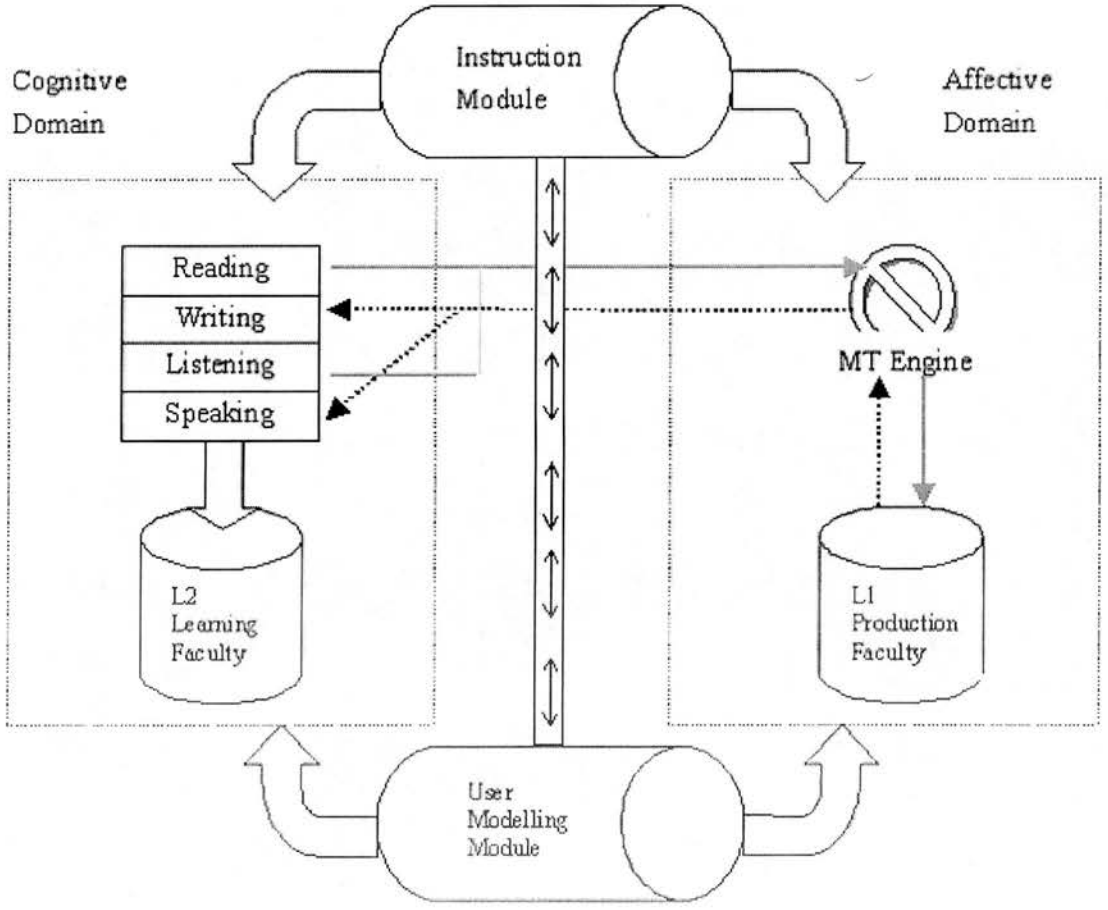


Figure9 : A balanced L2 learning model between cognitive domain and affective domain

The framework consists of at least an Instruction Module, a User Modelling Module, and an MT Engine, which mediates between the learner's L1 production faculty, L2 learning faculty, and the FL learning activities including reading, writing, listening and speaking.

- Instruction Module: This is the coordinator as well as the knowledge base of the framework. On one hand, it is preinstalled with expert knowledge in second language teaching which enables it to manipulate the language learning activities in specific models. On the other hand, the Instruction Module mediates between the MT Engine, the User Modelling Module, and the learning activities. It receives command from the User Modelling Module as to when to request service from the MT Engine for the student and how to integrate the translation help into the learning activities.
- User Modelling Module: This is the "care-taking" module of the framework. It is responsible for attending to the student's need for L1 support. There are a number of ways in which the student's activities can be monitored and difficulties spied. For example, in reading comprehension the user modelling can take the "word frequency approach" (Shei, 2001), which monitors the student's vocabulary development against a word frequency list extracted from a large corpus. The system can offer translation help automatically, drawing the user's attention to the sentences which contain a certain percentage of words lower on the frequency list than the student's calculated level of competence and thus likely to be unknown to him. In writing, on the other hand, where the student requests translation from L1 to FL, the monitoring can take the form of "service and record". That is, the system keeps a record of the sentences requested by the student to be translated into the FL. The history can be analysed at certain stages to find out the weaknesses of the student in the FL in terms of vocabulary, collocations, or syntax. More help features can be designed following the acquisition of this information.
- MT Engine: The MT system incorporated in this framework should be aware of the rules being applied while doing a specific translation, and be able to infer from a set of meta-rules the quality of the translation or the translation methods used,[49] i.e. whether it being a word-for-word translation, literal translation, or whatever categories the system attempts to distinguish. The MT can be a combination of the direct system, the transfer system, and the interlingua system,[50] the criterion being that the translation rules are transparent to the meta-rules and so the classification of translation methods is more tangible.
- Language learning activities: The kinds of language learning tasks that can be incorporated into this framework seem unlimited. The most important criterion of selection is: For which tasks can the use of L1 and translation achieve better results than if they are not used? For example, in beginning or

[49] See Newmark (1988:) for the discussion on translation methods.
[50] See for example Trujillo (1999: 5-6) for a description of these machineries.

lower-intermediate writing classes of English as a Foreign Language (EFL), students may benefit from MT output in certain ways, such as getting an easier start by initially translating their ideas in L1 into the FL through MT. Whereas in listening comprehension activities, MT is not much of a help since "spoken-language MT has not yet reached even the stage of real-time testing in non-laboratory settings" (Hutchins, 1999).

- L1 production faculty: This refers to the student's native-speaker competence in his first language. The student should have no problem producing fully grammatical and acceptable sentences in his L1. However, the MT system may require the student to use restricted language in order for the translation to be more acceptable.
- L2 learning faculty: This is the part of the language learner's competence responsible for interacting with the learning tasks, absorbing the FL input, shaping the learner's inter-language and producing FL output.[51]

9.3.5 Supporting Language Production

The proposed framework should be suitable for any FL learning environment, among which the scenario of foreign language writing is a prominent activity and will be considered here with the illustration of a worked example.

The student's task in this illustrative FL learning scenario involves the writing of a 'one-paragraph' composition for intermediate learners of English. The assigned topic is *Elections in Taiwan*, from which the student is to derive a topic sentence, like *Elections in Taiwan are noisy and expensive*, which incorporates the topic and the author's opinion (or controlling idea). Suppose the student is stuck with the first supporting sentences, which he can express in Chinese but not in English. The Chinese sentence in the student's mind is:

(1)

台灣	的	選舉	文化	充滿	了	謾罵	與	暴力
taiwan	*de*	*xuanju*	*wenhua*	*chongman*	*le*	*manma*	*yu*	*baoli*
TAIWAN	POS	ELECTION	CULTURE	FULL OF	ASP	INSULT	AND	VIOLENCE

An intelligent tutoring system could be programmed to detect the student's hesitation in composing, for example, where the student is neither typing nor doing research or checking online dictionaries. When hesitation time passes a threshold, the system

[51] The possible psycholinguistic interactions between L1, L2, and the learning faculties are beyond the scope of the proposed framework, hence are not touched upon.

offers translation help by asking the student to key in an intended sentence in L1. For the Chinese sentence in (1), for example, the system can offer a word-for-word translation like the one which follows the Pinyin[52] representation of the Chinese sentence closely in (1). Or it can offer a more refined version of literal translation as in (2), which is actually rendered by an existing online Chinese-to-English machine translation system:[53]

(2)
Election culture of Taiwan is full of abusing and violence.

The beginning to intermediate student can then learn the language through revising the MT output rather than being left to his own devices with his anxiety level rising. The revision procedure can be separately designed through human-machine interaction with the latter being preinstalled with a pedagogical grammar and electronic resources like language corpora and tools like dictionaries and concordancers to help the student decide on a final best version of the English translation. The system then resorts to the translation memory (TM) approach[54] by memorising the bilingual pair of the student's source sentence and the edited MT-generated target sentence. What's more, all the students' translation memories can be collected together through the local area network. In time, the machine will learn a large quantity of Chinese sentences whose translations are most demanded by EFL learners, as well as their refined English translations, which in turn makes the machine able to produce better translations upon invocation.

9.3.6 Supporting Language Comprehension

In reading comprehension the proposed framework helps the learner by offering translation for sentences incomprehensible to the learner. We will consider a case where the Chinese EFL student is reading the CNN Web page and encounters the sentence in (3) which he does not understand:

(3)
Three Japanese warships will depart for the Indian Ocean early Friday to provide non-combat support to the U.S.-led war on terrorism.

---

[52] The standard alphabetic system representing the sound of Mandarin Chinese.
[53] The URL for this MT site is: http://www.netat.net/big5/translation.htm.
[54] This is basically a kind of software which memorises every pair of source-target translation units (usually sentences) entered by the translator and fuzzily searches the database to see if it can find any identical or similar target sentence for each source sentence currently being translated. See Chapter 4 of Trujillo (1999) for a detailed explanation of translation memory systems.

The English-Chinese MT system we used to produce the text in (2) now translates the English sentence in (3) into the Chinese sentence in (4), with Pinyin and English word-for-word translation:

(4)

3 艘 日本 軍艦 星期五 的 早些 時候
san sao riben junjian xingqiwu de zaoxie shihou
THREE CLSF JAPAN WARSHIP FRIDAY POS EARLIER MOMENT
將 向 印度洋 出發
jiang xiang yinduyang chufa
WILL TOWARD INDIAN OCEAN SET OUT
關于 恐怖主義 提供 非 戰斗
guanyu kongbuzhuyi tigong fei zhandou
CONCERNING TERRORISM OFFER NON COMBAT
對 美國 領導 的 戰爭 的 支持。
dui meiguo lingdao de zhanzheng de zhichi
TOWARD AMERICA LEAD POS WAR POS SUPPORT

Although the Chinese translation is not entirely grammatical and acceptable, it is still quite informative and highly usable; i.e. it is helpful to the student's understanding of the original FL text. For the receptive FL learning tasks such as reading, as opposed to the productive ones such as writing, the system works in a different way in terms of student modelling. Now the focus is on modelling the student's progress rather then on attempting to offer idiomatic FL translations. One possibility is for the system to memorise the competence level or the "request level" of the user in terms of vocabulary or structure, and to anticipate the future requests he will made in the process of reading. The offer of translation to help the student to understand then becomes automatic.

To accomplish the above end the system needs to have an inbuilt word frequency list calculated from a large English corpus. The system lets the student do a diagnostic test, which decides on the vocabulary and grammatical levels of the student, as he first enters the system. Student modelling can take the form of vocabulary monitoring or the monitoring of other criteria such as sentence length. So for example, when a certain percentage of vocabulary appearing in a sentence which are lower in the frequency list than the student's computed level (and hence allegedly unknown to the student),[55] or when the length of the sentence passes a threshold, the

[55] See Shei (2001) for details of modelling the student's vocabulary level based on a frequency list.

system offers translation into L1 automatically. Alternatively, the system can record all the translation requests made by the user, and separate them from sentences for which translation is not asked for. The system can then build a statistical model for the request and another for the non-request sentences. Upon encountering a new sentence, the system runs it through the two models and decides which model applies and hence whether to offer translation or not. The system thus becomes an automatic translation helper during the student's learning processes, coming when needed and hiding when not needed.

9.3.7 Adaptation and Migration

Although the system, as contrived under the present framework, attempts to adapt to the learner's need for translation, the ultimate goal is for the learner to gain independence in the territory of the foreign language. Thus the system respects the learner's need to retreat to the use of L1, but insists on gradually migrating the learner to the use of FL in the language learning environment. The system does this by gradually withdrawing from providing full sentence translation, for example, in the FL writing scenario described in Section 5. Instead, the system now reverts to the offering of keywords or collocations related to the source sentences, which are nevertheless crucial for helping the learner to construct acceptable target sentences.

Shei & Pain (2001) discuss Chinese EFL learners' inabilities to produce grammatical and idiomatic sentences and suggest using MT generated sentence stems and collocations to help students construct acceptable English sentences. If the system could provide an appropriate sentence stem and crucial collocations for a prospective FL sentence, it would be of great help to the learner who is struggling to get the correct FL representations for the proposition in his head which could otherwise easily be materialised in his L1.

Take the Chinese sentence in (5) for example:

(5)

選舉人	爲了	拉攏	聲勢,
xuanjuren	*weile*	*latai*	*shengshi*
CANDIDATE	IN ORDER TO	PULL AND RAISE	REPUTE AND VIGOR

常	以	言語	猛烈	攻擊	或	詆毀	對手.
chang	*yi*	*yanyu*	*menglie*	*gongji*	*huo*	*dihui*	*duishou*
OFTEN	WITH	WORDS	VEHEMENT	ATTACK	OR	DEFAME	OPPONENT

Although the learner requests translation into English for this source sentence, the system nevertheless offers "translation help" instead of full translation in the first instance. The help is in the form of a sentence stem (or sentence builder) such as the ones shown below:

(6)
More often than not, X Y
X Y *in order to* Z

where X = NP, Y = VP and Z = VP, as well as suggestions of certain collocations, such as the ones that follow:

(7)

Verb-Noun Collocation	Adjective-Noun Collocation
enhance reputation	verbal abuse
attack rivals	vicious attack

Example sentences extracted from corpora illustrating the use of these sentence stems or collocations may also be included to further support the construction of an acceptable FL sentence. For the sample Chinese sentence in (5) which is to be translated into English by the learner, some possible supporting sentences from corpora are included below:

(8)

a.
They listen to their customers and make incremental improvements in mature systems of management and organisation to improve the service they offer people, by this means they **enhance** their **reputation** for quality and reliability.

b.
My Spanish cracked in the strain and I unleashed a torrent of **verbal abuse** whose tone he could not mistake.

c.
Within seconds of his return, Jessie launched a **vicious attack** on a table lamp, popping the bulb and wrecking the shade before charging into the lounge.

After the learner becomes comfortable in constructing sentences in this fashion, the next stage is to further encourage independence by offering the concordancer, the

dictionary and other electronic tools only in a passive mode; i.e. the student has to learn how to use these reference tools on his own initiative and gradually becomes a competent writer in the FL, rather than relying on machine translation inconclusively.

9.4 Summary and conclusion

This chapter introduced some newer varieties of CALL and discussed their relevance to translation teaching. In particular, the last section introduced a general framework for offering translation help to students learning a foreign language in a CALL environment. Although the success of a system under this framework depends to a certain degree on the advancement in machine translation, yet the most crucial functions prescribed by the framework can already be conducted under present MT technology. Hopefully, this illustrates how translation can be usefully incorporated into a CALL system.

Apart from serving the students at a department of translation, Type C programs as discussed above can also be considered for non-translation students to learn a foreign language by means of translating. What's more, as noted in the beginning of our discussion, the various types of CALL distinguished are actually features which can be integrated together. It is possible to have an intelligent, multimedia, corpus-based Web learning system, for example, if the educational setting so desires.

Chapter 10

Syllabus Design

This chapter considers the design of syllabuses incorporating elements of language teaching and translation teaching within a computer-assisted learning environment. Some researchers use the terms *syllabus* and *curriculum* interchangeably; for example, Crawford-Lange (1987) uses the term *curriculum design* in place of the more usual *syllabus design*. Others distinguish between syllabus design and curriculum development – for example, Rodgers (1989) defines the latter to be "a contextually enlarged view" of the former – a view adopted by this thesis, where *curriculum planning* refers to the organisation of courses for the whole department or school, and *syllabus design* means the planning of translation or language lessons at the individual classes level.

In principle the discussion about syllabuses should involve several aspects (or stages), from student needs analysis, to the development and implementation methods of a syllabus. Crawford-Lange (1987), for example, lists four "variables" to consider when structuring a curriculum – time, goals, learning strategies and materials, and evaluation strategies. Stern (1992: 41), on the other hand, suggests that curriculum management in second language teaching consists of three processes – development, implementation, and evaluation. However, because we deal with aspects of curriculum concerning needs analysis, teaching concepts, methods of implementation, etc. in previous or forthcoming chapters, in this chapter we are concerned with the result of the development stage (i.e. the content of the syllabus) only: we identify some prominent types of syllabuses in the history of second language teaching and the field of translation teaching. This information will form part of the basis for modelling our language-translation integrated syllabus.

Thus first we will look at the richer resources on syllabus design in the field of second language teaching. Then we examine the relatively meagre literature on translation syllabus. Finally, we ponder how syllabuses should be designed for trainee translators learning to translate into a second language.

10.1 Language teaching syllabus

An EFL (English as a Foreign Language) syllabus is defined by Sinclair & Renouf (1988) as

> ... a set of headings indicating items which have been selected ... to be covered in a particular part of the curriculum or in a course series. Its content is usually identified in terms of language elements and linguistic or behavioural skills.

Strevens (1987) further points out a commonly accepted doctrine in language teaching that "the effectiveness of learning can be significantly improved by manipulating the *content* and the *sequence* of what is taught". "Content" and "sequence" refer to the substance and arrangement of language learning materials, which, together with learning goals and teaching methods, etc. constitute the major concerns of syllabus design. Besides emphasising the usefulness of syllabuses, Strevens also notices the change of trends in syllabus design in the 1970s. He comments that "from *linguistic* to *situational* to *notional/functional* to *communicative* syllabuses, the sophistication has lately increased dramatically". Indeed, the concept of syllabus design has been through progressive evolution in the past thirty years. After the communicative syllabus, we have the task-based syllabus, the process syllabus, the lexical syllabus, etc. In this section, we explore types of language teaching syllabuses roughly in the order of their occurrences in history.

10.1.1 Structural syllabus

The structural (or the grammatical) syllabus corresponds to what Crawford-Lange (1987) terms the "Systems-Behavioral Design". The main feature of this type of syllabuses is that the contents for teaching are ordered grammatical units produced as a result of linguistic analyses of the target language. As Crawford-Lange observes, "the structure of the target language is first analyzed into discrete learning units. These are ordered, respecting any sequences inherent in the structure itself".

More often than not, the structural syllabus contains not only specifications of grammatical units but also the prescription and sequencing of vocabulary items. Yalden (1987) refers to this type of syllabus as the "Traditional Syllabus". According to Yalden, "this syllabus has generally consisted of two components: a list of linguistic structures (the 'grammar' to be taught) and a list of words (the lexicon to be taught)" (p. 19). Sinclair & Renouf (1988) note, however, that, for most syllabuses in this tradition, "pride of place is given to the grammar, and the vocabulary is clearly secondary".

The structural syllabus gradually lost its dominant position in language teaching since the 1970s, primarily due to its weaknesses in dealing with the functional aspect (i.e. language use) of language learning. Yalden (1987: 20), for example, notes that

"what has been missing from our language programs is a consideration of how we use language in everyday situations of all sorts". Thus various types of syllabuses addressing the issue of language use gradually came into fashion since the 1970s and took over the dominant place previously occupied by the structural syllabuses in language teaching.

10.1.2 Notional-functional syllabus

This type of syllabus has various names: notional syllabus (Wilkins, 1976), functional-notional syllabus (Crawford-Lange, 1987), or notional-functional syllabus (Markee, 1997). This syllabus was developed in response to a recognition of the weaknesses residing in the structural syllabus. Wilkins (1976: 10) points out:

> One of the major reasons for questioning the adequacy of grammatical syllabuses lies in the fact that even when we have described the grammatical (and lexical) meaning of a sentence we have not accounted for the way in which it is used as an utterance.

Function and notion are both concepts having to do with the use of language. Crawford-Lange (1987) explains, "Function is a matter of purpose.... Notion concerns the content of the purpose. For example, a person may ask (function) for a pen (notion)...".

According to Wilkins (1976: 2), the grammatical syllabus is a kind of synthetic syllabus, where "the learner's task is to re-synthesize the language that has been broken down into a large number of smaller pieces with the aim of making his learning easier". One of the complaints about this approach, according to Wilkins, is that "the motivation of learners is hard to sustain when success is measured in terms of the proportion of the grammatical system known" (ibid: 13). The notional syllabus, on the other hand, claims to take the analytical approach, which is "organized in terms of the purposes for which people are learning language and the kinds of language performance that are necessary to meet those purposes" (ibid.). In the analytical approach,

> Since we are inviting the learner ... to recognize the linguistic components of the language behaviour he is acquiring, we are in effect basing our approach on the learner's analytic capacities" (ibid: 14)

However, when one looks at the components of a notional syllabus as proposed by Wilkins, they are not very different from the units specified in a structural syllabus. The specifications of semantico-grammatical categories in Wilkins (1976: 25) such as

time, quantity, space, etc., cannot really guarantee that the language learner will associate these with the use of language more readily than they do with grammatical categories. What's more, the categories of modal meaning (ibid: 38) look almost exactly like the section on modal auxiliaries in a structural syllabus. Indeed, as Markee (1997: 17) observes,

> Nonetheless, notions and functions are still linguistic units of analysis. Using preselected *linguistic* units and linguistic criteria to select, grade, and sequence pedagogical content leads us back to synthetic syllabus design solutions.

Despite the above criticism, the notional-functional syllabus has enjoyed some success and its influence is far-reaching, probably because of its sound and appealing emphasis on language application in syllabus design. Markee (1997: 17) comments,

> Functional syllabuses in particular have enjoyed great popularity; indeed, by the late 1970s and early 1980s, functional materials had spread all over the world.

Overall, however, the notional-functional syllabus as proposed by Wilkins (1976) do not seem to offer a framework comprehensive enough to cover both content and use of language, or a viable way for coordinating both. Nevertheless, the attention of EFL practitioners was successfully drawn to the functional aspect of language teaching, and a foundation was laid for the communicative syllabus to be proposed later.

10.1.3 Communicative syllabus

Yalden (1987: 86-87) proposes ten possible components for the communicative syllabus:

1. purposes
2. setting
3. role
4. communicative events
5. language functions
6. notions
7. discourse and rhetorical skills
8. variety
9. grammatical content
10. lexical content

An ideal communicative syllabus should then incorporate elements of the structural syllabus (9 and 10), as well as those of the notional-functional syllabus (5 and 6). Contrary to the common belief that communicative language teaching disregards grammar partially or entirely, the ideal communicative syllabus actually looks upon grammar as an essential part in an EFL syllabus. Stern (1992) notices this comprehensive nature of the communicative approach and comments:

> Although it has incorporated many of the characteristics of earlier language teaching innovations, communicative language teaching has avoided the narrowness and dogmatism of the method concept and covers a wider range of components.

The communicative syllabus in its ideal form seems very difficult to design and implement, probably because there is no theory to back up the analysis of the components of communication on which a pedagogical model could be built. Sinclair & Renouf (1988) note:

> There is now a large group of notional, functional and communicative syllabuses which reflects a different theory of language, deriving not so much from traditional linguistics as from theories of discourse based on speech act philosophy.

They further point out that "there is no comprehensive theory of language in these terms available as yet, so such syllabuses rest on shaky ground" (ibid).

Interestingly, the appealing nature of the communicative approach and the difficulty of defining a good communicative syllabus seem to lead to the creation of a middle ground, as exemplified by the contemporary EFL textbooks, which seem to attempt to incorporate a "hidden syllabus" in a series of richly stocked textbooks. These textbooks do not profess to conform to any types of syllabus, but they do endeavour to include many components of a communicative syllabus as proposed by Yalden (1987). For example, a chapter in Harris & Rowe (1997: v) has the following structure:

Chapter 1

TOPICS
Politics
Meeting people
Movies

GRAMMAR
Present perfect continuous

Verb + gerund
Gerund as subject

FUNCTIONS
Talking about things people have been doing
Expressing likes and dislikes
Giving opinions
Telling story
Predicting

There are two things to notice from the above: First, a chapter summary like this in a contemporary EFL textbook looks remarkably like the specifications of grammar or notions etc. in a syllabus. Secondly, the contents of a lessons seem to cover the main aspects of a communicative syllabus: notions, functions, and grammar. If we look at the lesson itself, there are also components of vocabulary and activities which draw attention to settings, discourse skills, etc., which conform to the requirements of a communicative syllabus. Similarly, another textbook, the *New Cambridge English Course*, lists six dimensions of learning for each lesson: grammar, phonology, functions, topics and notions, skills, vocabulary (see McDonough & Shaw, 1993).

Sinclair & Renouf (1988) note that:

> ... there is for language teachers in state schools and private organizations increasingly no distinction between syllabus, methodology and coursebook. All are blended in an officially blessed publication...

Indeed, for colleges and universities in Taiwan as well, although EFL courses are compulsory for the freshmen, there is no compulsory syllabus for teachers. Many EFL instructors simply adopt series of textbooks like Harris & Rowe (1997) and do not worry about the specification or implementation of a syllabus at all. The textbook itself is a syllabus.

This type of "coursebook-based syllabus", however, is severely criticised by Sinclair & Renouf (1988), who claim that "a syllabus which is dependent on a particular coursebook is a degenerate syllabus, not very much different from the table of contents". For Sinclair & Renouf, the pre-specification of a language teaching syllabus is very important. Hence the lexical syllabus they propose embodies their attempt to conform to this principle, as we shall see in section 10.1.5.

10.1.4 Task-based syllabus

The task-based syllabus followed on the heel of the communicative approach and became popular after the mid-1970s. As its name implies, the task-based syllabus

comprises a set of tasks which involve using the target language to do something. The task-based syllabus generally does not refer to any linguistic or functional principles for organising its units. According to Nunan (1989: 10), a task is

> a piece of classroom work which involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is principally focused on meaning rather than form.

Thus the task-based approach shares the same language learning philosophy with communicative language teaching, both of which emphasise meaningful use of language during communication. A notable difference is that the task-based approach does not specify its syllabus in functional or structural terms. Nunan (2001) explains:

> Task-based syllabuses represent a particular realization of communicative language teaching. Instead of beginning the design process with lists of grammatical, functional-notional, and other items, the designer conducts a needs analysis which yields a list of the target tasks that the targeted learners will need to carry out in the 'real-world' outside the classroom.

Long & Crookes (1992) recognise three types of task-based syllabus: (1) the procedural syllabus, (2) the process syllabus, and (3) the task syllabus. All of these are said to be a kind of analytic syllabus, since they present language as wholes to the learner and the learner has to resort to his linguistic capacity to analyse the language he encounters.[56] This coincides with what Sinclair & Renouf's (1988) recognise about the task-based syllabus, who point out that "it is held that, properly designed, such a syllabus will cover a sufficient range of vocabulary, grammar, notions, functions and skills". Another way of saying this is that all grammatical and functional materials to be learned by the student are subsumed in the tasks and need not be explicitly analysed or stated in the syllabus. The syllabus is just a collection of unanalysed tasks.

The task-based syllabus offers a kind of eclecticism to language teachers. According to Prabhu (1987), the procedural syllabus, a type of task-based syllabus, does not have a fixed set of descriptions. It is simply a memorandum passed from one teacher to another in order for the teachers to share the experience of conducting task-based language lessons. Prabhu says:

---

[56] Recall that for the synthetic syllabus (e.g. structural syllabuses), the student is presented with analysed parts of the target language, and his task is to re-synthesise the parts into meaningful wholes.

> The teachers who drew on the collection in the teaching of later project classes altered the sequencing of tasks at various points, modified the content of some of the tasks in order to raise or lower the difficulty level... (p. 178)

A teacher following the methodology of communicative language teaching, then, can choose a textbook like Harris & Rowe (1997: v) and follow the 'syllabus = coursebook' approach, which yields little flexibility; or take the task-based approach and organise her language lessons based on a collection of tasks, which offers maximum flexibility.

10.1.5 Lexical syllabus

An important trait of a lexical syllabus, according to McDonough & Shaw (1993: 52) is for vocabulary to be taught in context, especially in an authentic context. For example, the COBUILD dictionary is claimed to have been compiled on the basis of a large corpus: the 450 million word Bank of English (n.d.). Each entry in the dictionary contains example sentences taken from the corpus illustrating the usage of the word. Such an approach to vocabulary is based on authentic context.

The lexical syllabus proposed by Sinclair & Renouf (1988) is also based on a corpus: in this case the 7.3 million word Birmingham Corpus. Sinclair & Renouf firmly believe that a language syllabus must exist independent of the textbook. They claim:

> For a syllabus to have an important role in education, it should either pre-exist or be devised independently of other elements like course materials, methodology, and assessment.

Sinclair & Renouf consider three factors in deciding on the content of the lexical syllabus: what counts as a word? which words to teach? what to teach about words?

In discussing what constitutes a word, Sinclair & Renouf argues that the conventional view about word forms is not appropriate for language teaching. Conventionally, the concept of 'word' includes a base form together with all the other inflectional forms. So that, for example, *certain* and *certainly* would be treated as the same 'word' (or 'lemma' in computational terms) in the traditional way. However, Sinclair & Renouf point out that different word forms like these have very different senses and should be treated as separate 'words' in a language teaching syllabus. They say:

> ... the individual word forms are so different from each other in their primary meanings and central patterns of behaviour ... that they are essentially different 'words', and really warrant separate treatment in a language course.

Other words they suggest in the same vein are: *easy* versus *easily*, *near* versus *nearly*, *real* versus *really*, etc.

Once the concept of 'word' is clarified, the other things to consider in contriving a lexical syllabus are: which words to select and what about these words to teach. According to Sinclair & Renouf, the main focus of study for any learner of English should be:

a) the commonest word forms in the language
b) their central patterns of usage
c) the combinations which they typically form

With regards to word selection, Sinclair & Renouf propose a balance between "natural usage" and "utility". Natural usage is inferred from a corpus and the words selected are the words most frequently used by native speakers; for example, the first 800 or so items in the word frequency list. Besides these common words, according to Sinclair & Renouf, the word list on a lexical syllabus should also include some high-utility words for language learners which are perhaps not so frequently used by native speakers; for example, words relating to "domestic reality", "physical sensations", "personal emotions", etc.

As for things about these words to teach after the selection has been made, Sinclair & Renouf suggest using corpora again to extract the most common usages of words. For example, most native speakers (and the syllabus designer) would intuitively think that the primary usage of the word *see* is the act of seeing through one's eyes. However, as Sinclair & Renouf point out, "Textual evidence shows us that the first and second most frequent uses of *see* are those found in the familiar phrases *you see* and *I see*". Sinclair & Renouf consider the statistical evidence derived from a corpus a helpful means for distinguishing the commoner patterns of usage from the less common ones.

Another thing which Sinclair & Renouf consider essential in teaching vocabulary is the combination of words (i.e. collocation). They note both lexical collocations (e.g. *happy marriage* or *accidental death*) and grammatical collocations (e.g. *each hour/day/week* and *kind/part/sort of*), which are also a focal point of this thesis (see Chapter Five).

10.2 Translation syllabus

Research on translation syllabuses has been scarce. As Vermeer (1998) observes, "translation teaching and learning have ... traditionally relied on the old methods of language teaching and learning". Thus a traditional syllabus for translation should bear strong resemblance to a traditional (i.e. structural) language teaching syllabus.

Just like the change brought to language teaching syllabuses by the communicative approach, however, the idea for teaching translation also changed as people increasingly see translation in a cultural context rather than simple transfer of linguistic signs. Vermeer's words are again relevant:

> ... viewed as an act of intercultural communication rather than a skill in transferring minimal linguistic units across language boundaries, translation could no longer be taught/learnt on the basis of linguistic exercises.

Vermeer discusses this new concept in teaching translation under the rubric of "the functional approach", the same term used for the functional/communicative 'revolution' in the field of language teaching in the 1970s.

The next stage of revolution in translation teaching should be the one associated with the translator's workstation (TW) and machine translation (MT). Melby (1994), for example, distinguishes three components of a translator workstation: the hardware, the operating environment, and the application software, which can offer a foundation for designing TW-based translation syllabuses. Wältermann (1994), on the other hand, describes a translation curriculum based on machine translation and conducted in the environment of a translator's workstation. Thus in the following, we look at the three categories of translation syllabus, from the traditional to the most contemporary.[57]

10.2.1 Linguistic syllabus

The linguistic syllabus is used here to refer to a syllabus for teaching translation which draws on a contrastive analysis of the two languages in question, and sets out learning objectives and lesson plans based on the observations made and translation skills inferred from the analysis. Vermeer (1998) notes the type of linguistic equivalence rules like "translate German adverbs by a Spanish final verb + *que* construction and vice versa...", which constitutes a translational skill, which "is taught over and over again, up to eight times per diploma course".

[57] Terms for the three kinds of translation syllabus (i.e. linguistic, cultural and technical syllabus) are proposed by the present author and are not regular terms in the literature.

Zhang et al. (1993) offer an English-to-Chinese translation coursework which is typical of a linguistic syllabus for translation teaching. The book has the following (brief) table of contents:

Chapter One: Introduction to the history of translation in China
Chapter Two: The standards and processes of translation
Chapter Three: Contrasts between the English and Chinese languages
Chapter Four: Common methods and skills for English-to-Chinese translation (I)
Chapter Five: Common methods and skills for English-to-Chinese translation (II)
Chapter Six: Common methods and skills for English-to-Chinese translation (III)

The main contents of the book reside in Chapters Four to Six, which explain translational skills mostly like that described by Vermeer, for example, the principles for translating English passive sentences, the techniques of word addition, deletion, and repetition, the skills for translating long English sentences etc.

The linguistic syllabus provides the learner with a good foundation for the bottom-up approach to translation, but dwelling on grammar and vocabulary alone is not adequate for producing high-quality translation. A more global, or top-down approach is indispensable, which includes notably the consideration of culture.

10.2.2 Cultural syllabus

Newmark (1991: 130) proposes that "the success of any translation course must depend 65% on the personality of the teacher...".[58] The personality of the teacher, Newmark goes on to suggest, includes three things: personal qualities, professional qualities, and general knowledge of culture (ibid.: 131). More specifically, "the translation teacher has to be not only a solid classroom teacher, but a person of wide cultural background" (ibid: 138). Likewise, Nida (1997) points out that although current translator training programs have a balanced focus on theory and practice, "what is often lacking, however, is the broader cultural background that results from prior university training".

Why is the understanding of culture important to a translator? Krouglov (1996) discusses the importance of considering social and cultural factors in teaching translation and interpreting. Krouglov gives some vivid examples to illustrate his point. The first example is a Russian guide greeting a group of touring students with the Russian equivalent of "Dear students". This literal translation is exotic and inappropriate when considering the target language (i.e. English) norms – "Hello"

[58] The other percentages are: "20% on the course design and 15% on the course materials".

would have been used in the English culture for such a situation. Similarly, Krouglov explains that, in Russian, the person answering a phone call asking for himself would say the Russian equivalent of "Listening". But to translate the formulaic expression into English literally is "incongruous", in Krouglov's own words. In translating from Russian to English, the phrase "Listening" in telephone conversation "must be replaced by the communicative equivalent 'Speaking'". According to these observations, Krouglov points out the principle of his course design:

> My primary concern at the first year level is therefore to teach the students not to mechanically substitute lexical items but to make them aware of the social and cultural significance of various situations...

Despite the necessity of having socio-cultural ingredients integrated in a translation syllabus, as far as the author is aware, however, there is as yet no established work proposing the details of or the guidelines for constructing such a syllabus. The cultural syllabus proposed here should include the linguistic syllabus, just as the communicative syllabus for language teaching subsumes the grammatical syllabus. That is, a cultural syllabus must be built upon or integrated with a linguistic syllabus. Newmark (1991) supports this view indirectly and gives some useful hints when he talks about the kinds of notes a student should take in a translation class. Newmark says, "I encourage them [i.e. the students] to bring notebooks, and compile (a) grammatical, (b) lexical, (c) cultural notes under separate heads... " (p. 131). This seems to indicate that a standard translation course in Newmark's design would include notably grammatical, lexical and cultural aspects. A cultural syllabus, that is to say, will approach the issue of translation from these three perspectives and include materials which are able to contrast grammatical, lexical and cultural differences between the source and target texts.

More specific possible elements in a translation cultural syllabus are proposed in Newmark (1991:137), who says:

> SL and TL cultural backgrounds are essential subjects in a university translation course. ... Their basis is solid economic geography, a conspectus of a country's industrial services and agricultural centres related to demography, social and political factors, as well as to scientific and cultural achievements. This is followed by a description of national institutions: central and local government, corporations, the churches, utilities, education, arts centres and the current philosophies which guide and assail them.

As we can see, Newmark's above specifications could be a very good start for constructing a cultural syllabus for teaching translation.

10.2.3 Technological syllabus

This category is proposed to describe the translation course which aims to follow the newest trend in translation in the context of modern technology. According to Boettcher (2001), contemporary information technology not only adds to the power of education, but even prompts a reconsideration of the entire educational system, including the concept of syllabus design. O'Hagan (1996) discusses the "teletranslation" service in a futuristic perspective, which is an integrated translation service combining the technologies of telecommunication, machine translation, and speech recognition. O'Hagan's discussion should provoke the reconsideration of a translation curriculum and the creation of a more innovative translation syllabus in order to meet the demands of language and information technologies. This section tentatively calls this kind of translation syllabus a "technological syllabus".

Standing in the core of a technological syllabus is the principle of using machine translation and machine-aided translation tools in teaching translation. Wältermann (1994), for example, discusses translation syllabuses with integrated machine translation components. The motivation for introducing such a kind of syllabus, in Wältermann's words, is the recognition that "future translators must be brought up to a level of sophistication when whey are routinely able to use state-of-the-art computational environment".

Wältermann introduces a two-course sequence for teaching scientific and technical translation, which covers the following modules:

Scientific-Technical Translations: Course 1

A. Historic and Linguistic Background to Technical and Machine Translation
B. Introduction to the Translator's WorkStation

Scientific-Technical Translations: Course 2

A. Setting up the translation project
B. Performing the translation project: Pre-translation Stage
C. Document production: Translation Stage
D. Post editing of Translation Document
E. Maintenance of MT Systems
F. Discussion and Evaluation

The working environment of the above sequence of courses is a "pedagogical translator's workstation" described in Wältermann (1994), which comprises machine

translation as well as machine-aided translation features. Wältermann explains that "besides giving the students the linguistic and cultural skills required of translators, they are being trained in maintaining and updating MT systems". This also reminds us that a technology-based syllabus does not ignore the linguistic and cultural aspects of translation learning, but subsumes them.

10.3 The combined syllabus

This section discusses the principles of syllabus design for translation courses aiming to enhance students' competence in both language (i.e. L2 or the target language) and translation. As such a syllabus involves the learning of both language and translation, it is useful to draw from the above discussions on syllabus design for both language teaching and translation. It is also essential for the proposed syllabus to incorporate elements of CALL and materials in relation to translation technologies.

10.3.1 Category of course

Teaching translation as a stand-alone practice is a relatively new enterprise. Translation teaching has been and is still being confused with language teaching. Newmark (1991:137) says, "Teaching translation, like translating, is a new profession, and ... it needs distinguishing from teaching translation within language teaching...". In Figure 10-1, I distinguish four categories along the language teaching-to-translation teaching continuum. Newmark is concerned with separating category 4 from category 2; whereas the focus of this thesis is on the third category, integrating translation teaching with the learning of L2, the discussion on which has still been sparse in the field of translation teaching.

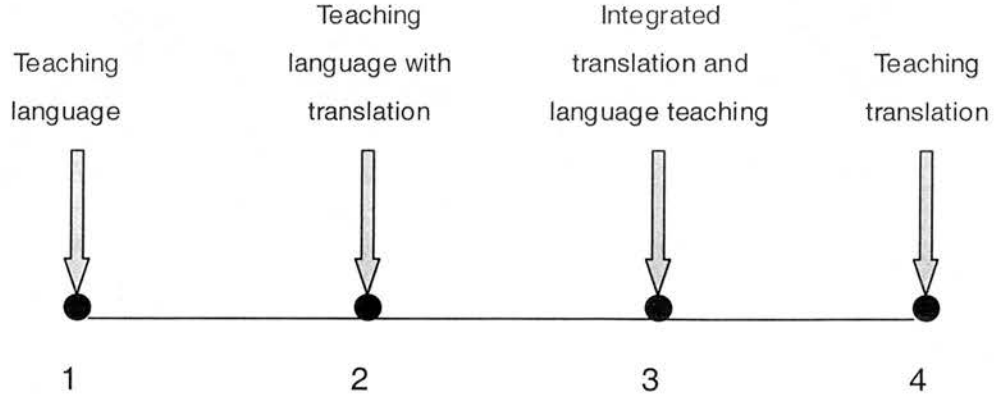


Figure 10-1: The Language and Translation Teaching Continuum

First of all, the segment between category 1and category 2 itself represents a band of language teaching methodologies with two extremes – on one end, translation is extensively used for language learning; on the other, translation is prohibited. Newmark (1991:50) notes the issue by saying that "the place of translation in foreign language teaching will always be dependent on the role that the learner and the teacher assign to the native language in the learning process". He refers to the two extremes about using translation in foreign language teaching by the names of, first, the Bilingual Method (where translation is maximally used) and, second, the Direct Method (where translation is not used). The notorious Grammar-Translation method, now largely abandoned in its canonical form, is also a category 2 practice. No matter the focus of teaching leans towards category 1 or category 2, however, the overall emphasis is on acquisition of the foreign language. Translation, is this context, is only one of the many means that serve foreign language teaching.

The point Newmark makes to distinguish category 4 from category 2 is well-motivated. Some researchers still mix the discussions of these two categories inappropriately, perhaps subsuming category 4 under category 2 in a "predestined" way. For example, Krouglov (1996) discusses the importance of socio-cultural aspects in teaching translation and interpreting. This can be seen from Krouglov's main concluding remark, which claims that "a deep understanding of the ways in which social and cultural features are combined in a language is indispensable in teaching translation and interpreting". Curiously, although Krouglov has not included any discussion on foreign language teaching at all in this article, or on the relationship between translation and language learning, for that matter, he nevertheless mentions foreign language teaching in several places in such a way that it looks as if language teaching were an ultimate and inevitable framework only within which the teaching of translation could be discussed. For example, in the beginning of the article, Krouglov says,

> In this article I evaluate the use of translation and interpreting in teaching a foreign language while focusing on problems of social and cultural differences which instructors face in everyday teaching.

This seems to lead the reader to think that the forthcoming discussions would be focusing, at least in part, on the role translation and interpreting can play in foreign language teaching, though in fact, nothing like this is mentioned in the article, which dwells almost exclusively on the implications of socio-cultural differences for teaching translation. [59] Also, the sentence leaves one wondering who the

[59] Though there is a confusing section title "Teaching translation at the initial stages of language learning" in Krouglov (1996), which seems to promise a discussion on language learning from, for

"instructors" facing "problems of social and cultural differences" daily are. On one hand, it should refer to foreign language teachers because the higher structure this sentence refers to is the language teaching part. On the other hand, it should refer to translation teachers, however, since these are the more legitimate persons dealing with the "problems of social and cultural differences" in translation. All in all, this illustrates the problem of not distinguishing between categories properly along the language and translation teaching continuum.

The focus of this thesis is on category 3, which has rarely been discussed in the literature of translation teaching (or language teaching). As the emphasis of category 2 teaching is on language (i.e. leaning towards category 1), so the emphasis of category 3 teaching is on translation (i.e. leaning towards category 4), but that is where the analogy ends. Category 3 is not, as the analogy would go on to predict, "using language teaching as a means for acquiring translation skills". Instead, category 3 teaching is the integration of foreign language teaching and translation teaching in a translation curriculum. The principal goal of this type of teaching is for the student to produce grammatically correct and idiomatic translation in the target language which is a foreign language to him. The principal methodology for achieving this goal, as proposed in this thesis, is for the student to learn to use tools in a translator's workstation to acquire the competence and skills to approximate native-like idiomaticity in producing translation.

Category 3 differs from category 2 in that the goal of category 3 is translation competence built on foreign language competence; whereas the goal of category 2 is mastery of the target language with translation as a means for achieving this goal and with no special reference to students' competence in translation. Category 4 differs from category 3, on the other hand, in that category 4 does not lay special emphasis on the acquisition of the foreign language as a target language in translation; whereas category 3 emphasises the mastery of both L2 competence and translation skills, both of which may help each other in a spiral manner (i.e. mastering the L2 helps build up translation competence; pursuing excellence in translation motivates the quest for better L2 abilities).

10.3.2 Designing principle

Graves (2000) suggests that the processes of course design should be represented as a cyclic flowchart rather than a linear list. In a traditional model of syllabus design, the processes of course development are largely linear and deterministic – for

---

example, a psycholinguistic perspective, again no theories remotely related to applied linguistics are mentioned and the discussions are purely translational.

example, the processes of development, implementation and evaluation of a curriculum introduced in Stern (1992: 41-48). Graves' model, in contrast, emphasises that 'there is no hierarchy in the processes and no sequence in their accomplishment' (p. 3). Graves takes a 'system approach' to syllabus design, where the processes of course development are interrelated, changing one component will influence the others. This model is particularly suitable to a computer assisted language/translation teaching syllabus incorporating educational software, since the integration of such software introduces more unpredictable variables into the syllabus which need to be monitored and modified closely for the ultimate success of the syllabus.

The incorporation of the use of software into a traditional syllabus requires specially contrived methodologies. Hubbard (1996) offers a framework which can be useful to us, especially its implementation module. Roughly speaking, Hubbard's implementation model comprises three layers: the hardware component (the accessibility issues), the learner components (the use of courseware, the preparatory and the follow-up activities), and the teacher components (teacher control, including authoring of the software, classroom management, student records, etc.). Based on this model, a model for integrating computer software into a General Translation curriculum can be proposed, as in Figure 10-2.

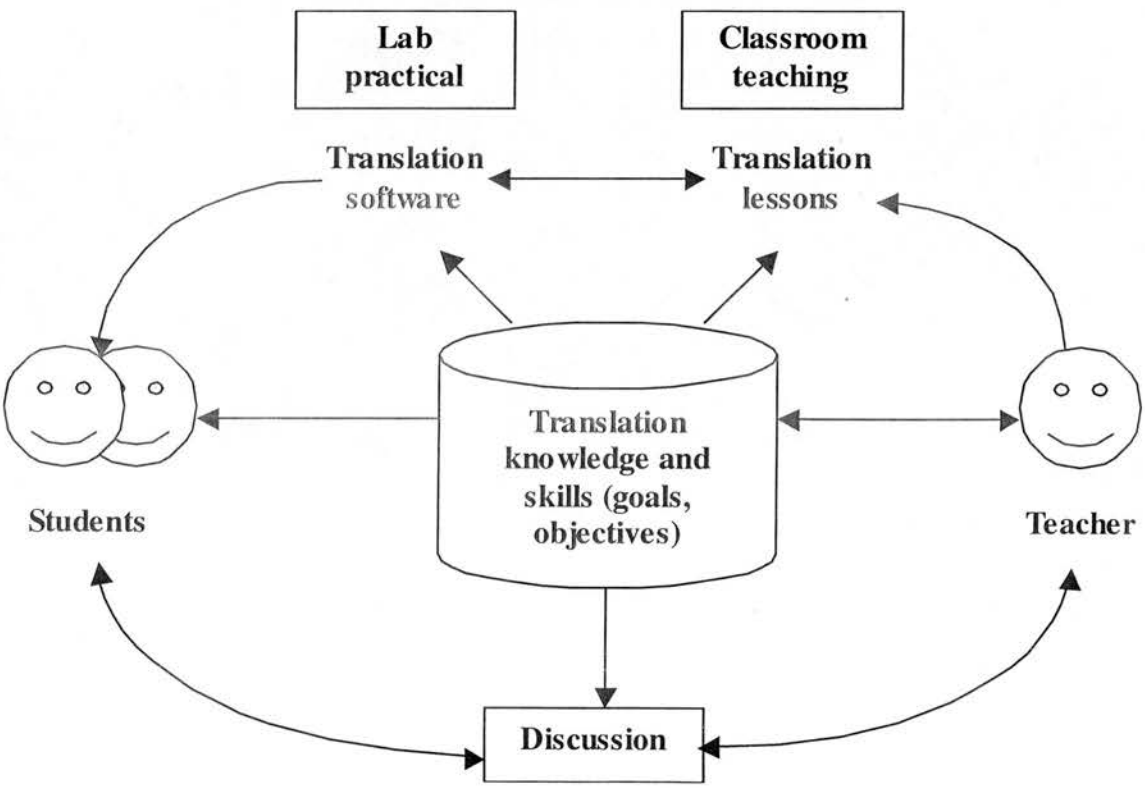


Figure 10-2: An implementation model of an L2-translation combined syllabus

The model emphasises the assumption that processes in syllabus designing are cyclic – they inform and influence the refinement of each other. Standing in the middle are the initial goals and objectives of the syllabus represented by pedagogical statements concerning the fulfilment of translation knowledge and skills, which direct all the activities in the cycle as the one-way arrows indicate. The goals and objectives are revisable by the teacher, however, as the two-way arrow shows. The translation lessons and the translation software modules constitute the main class activities which are mutually informative and modifiable. The teacher initiates the translation lessons and provides the translation related software, both of which subsume learning activities in the classroom or lab. For example, the lessons may include reading theories of contrastive linguistics and rules about translation. The lab sessions, on the other hand, consist of the application of such rules and possibly other supporting courseware. Based on the follow-up discussions between the students and the teacher, both the translation lessons and the translation software can be modified by the instructor for the future sessions.

10.4 Evaluation

It is difficult to evaluate the success or not of a language teaching (or translation) syllabus in relatively short terms. Post-session questionnaire is a humble but feasible way to probe into the effect of the course on students. Questionnaires may ask students to evaluate the program, the instructor, or themselves. In whichever case, the answers are useful information and constitute an initial estimate of the achievement (or non-achievement) of the syllabus.

A case in point is the course entitled "English and the computer" instructed by the author at National Tsing Hua University (NTHU), Taiwan, during the fall semester of 2002, which adopted a kind of syllabus lying between Type 1 and Type 2 as discussed in 10.3.1 (Figure 10-1). This was a general course in EFL (Type 1) designed as a kind of task-based syllabus (Nunan 1989). Although the syllabus covered a wide range of activities like reading and writing, it also involved some translation (Type 2). Thus this was an initial approximation to the combined approach to teaching language and translation. The weekly schedule of the course was as follows:

Week 1: Introduction
Week 2: Concordancing and language learning
Week 3: Concordancing and Vocabulary in Reading
Week 4: Concordancing and collocation

Week 5: Lexical phrases
Week 6: Phrasal constraints
Week 7: Sentence builders
Week 8: Making personal homepage in L2
Week 9: Post-editing MT output
Week 10: Pre-editing MT input
Week 11: MT and L2 reading
Week 12: MT and L2 composition
Week 13: Translation tools
Week 14: Translation Task
Week 15: Tagging and composition
Week 16: Bilingual concordancing
Week 17: More on bilingual concordancing

Thus weeks 9-14 involved the use of translation for L2 learning. For example, the partial text written by the instructor explaining the task of week 10 for students to complete looks like the following[60]:

[60] Illustration by examples of source texts and MT outputs contained in the original teaching documents, is omitted here.

Pre-editing Input to Machine Translation

Goal of Task: Practice editing Chinese input to an MT system and, from observing the MT output in English and judging its grammaticality and correctness in word selection, etc., decide on how to alter the Chinese text in order to get better MT output.

Purpose of Task: To familiarise yourself with the recent MT technology, make you aware of how MT can help you learn English perhaps, and let you test ideas about the two languages in question by playing with automatic translations between them.

Output of Task: You should record all the stages of your modification to the original Chinese supplied and the MT output thereof. You should also mark down your observations, your hypotheses about the nature of the MT errors observed, and your proposals about favourable changes to the input text at each stage, and, at the subsequent stage, whether the MT system has responded to your hypotheses correctly. Finally, you should summarise your investigation at the end of your report giving an overall impression of the capacities of the MT, how it might help your learn English, and what your have learned about Chinese and English from this task.

Figure 10-3: An excerpt of online handout for a translation task

It should be emphasised that this course was only partially related to translation and L2 learning, as it was a general English module and not a translation one,[61] and so the discussion on its evaluation is of a tentative nature wishing to shed some light on a truly L2 and translation combined approach.

Towards the end of the semester, an evaluation sheet in the form of a questionnaire was officially issued by the university, NTHU, for students to complete for each of the modules they took. While this was not especially designed by the author (i.e. the instructor of the course) with the target type of course in mind, it is nevertheless hoped that the questionnaire can show some of the effects of the combined approach to teaching language and translation, especially in comparison to

[61] Thus its devotion to a fully translation-based methodology could not have been justified within the current academic setting, where L2 teaching is generally dominated by the communicative language teaching approach.

a more traditional type of L2 learning module. The fifteen questions on the questionnaires are translated in Table 10-1 (the original questionnaire was in Chinese), together with the result of student marking. In Table 10-1, Class A represents the "English and the computer" class mentioned above, and Class B represents another general English class taught by the author – only this one did not involve the use of translation. Both A and B are undergraduate general English courses taken by the freshman entering NTHU in fall 2002. In contrast to the task-based approach for Class A whose classes were primarily conducted in the computer laboratory, the author used a more traditional approach for Class B, where the syllabus was based mostly on reading comprehension and vocabulary acquisition and their classes were conducted mainly in an ordinary classroom based on hardcopy textbooks.

Questions	Class	Very much agree	Agree	Mostly agree	Disagree	Very much disagree	Total
1. The materials and contents of this course are well organised and appropriate, the contents corresponding to the goals satisfactorily.	A	2 (11.8%)	12 (70.6%)	3 (17.6%)	0 (0%)	0 (0%)	17 (100%)
	B	0 (0%)	5 (26.3%)	10 (52.6%)	4 (21.1%)	0 (0%)	19 (100%)
2. The homework and exercises match the teaching content well, and the quantity and difficulty level are just right.	A	2 (11.8%)	12 (70.6%)	3 (17.6%)	0 (0%)	0 (0%)	17 (100%)
	B	1 (5.3%)	5 (26.3%)	10 (52.6%)	3 (15.8%)	0 (0%)	19 (100%)
3. The instructor teaches in a serious manner, and is able to inspire the students' studying and thinking.	A	3 (17.6%)	11 (64.7%)	3 (17.6%)	0 (0%)	0 (0%)	17 (100%)
	B	6 (31.6%)	10 (52.6%)	2 (10.5%)	0 (0%)	1 (5.3%)	19 (100%)
4. The instructor interacts well with the students, encouraging them to ask questions and to discuss, creating an environment conducive to learning.	A	1 (5.9%)	14 (82.4%)	2 (11.8%)	0 (0%)	0 (0%)	17 (100%)
	B	4 (21.1%)	9 (47.4%)	5 (26.3%)	0 (0%)	1 (5.3%)	19 (100%)
5. This course helps the students' intellectual growth, enhancing students' abilities in the subject area.	A	2 (11.8%)	11 (64.7%)	4 (23.5%)	0 (0%)	0 (0%)	17 (100%)
	B	2 (10.5%)	8 (42.1%)	6 (31.6%)	2 (10.5%)	1 (5.3%)	19 (100%)
6. The evaluation method adopted by the instructor reflects the objectives of teaching and can test the result of learning.	A	1 (5.9%)	13 (76.5%)	3 (17.6%)	0 (0%)	0 (0%)	17 (100%)
	B	4 (21.1%)	4 (21.1%)	9 (47.4%)	1 (5.3%)	1 (5.3%)	19 (100%)
7. The instructor very rarely skips class, changes schedules, comes late or leaves early.	A	6 (35.3%)	9 (52.9%)	2 (11.8%)	0 (0%)	0 (0%)	17 (100%)
	B	5 (26.3%)	7 (36.8%)	6 (31.6%)	1 (5.3%)	0 (0%)	19 (100%)

8. The instructor has expertised knowledge in the subject area, and can explain and express ideas very clearly.	A	8 (47.1%)	8 (47.1%)	1 (5.9%)	0 (0%)	0 (0%)	17 (100%)
	B	7 (36.8%)	6 (31.6%)	5 (26.3%)	1 (5.3%)	0 (0%)	19 (100%)
9. The instructor takes students' requirements into consideration, and can adjust to student needs.	A	6 (35.3%)	11 (64.7%)	0 (0%)	0 (0%)	0 (0%)	17 (100%)
	B	4 (21.1%)	8 (42.1%)	6 (31.6%)	1 (5.3%)	0 (0%)	19 (100%)
10. The instructor gives adequate, constructive feedback to student enquiries and homework.	A	5 (29.4%)	10 (58.8%)	2 (11.8%)	0 (0%)	0 (0%)	17 (100%)
	B	5 (26.3%)	10 (52.6%)	3 (15.8%)	1 (5.3%)	0 (0%)	19 (100%)
11. The instructor is well-prepared for classes.	A	7 (29.4%)	9 (52.9%)	1 (5.9%)	0 (0%)	0 (0%)	17 (100%)
	B	3 (15.8%)	13 (68.4%)	2 (10.5%)	1 (5.3%)	0 (0%)	19 (100%)
12. The quantity of teaching for each class session is appropriate.	A	4 (23.5%)	10 (58.8%)	3 (17.6%)	0 (0%)	0 (0%)	17 (100%)
	B	1 (5.3%)	8 (42.1%)	7 (36.8%)	2 (10.5%)	1 (5.3%)	19 (100%)
13. The instructor evaluates students fairly and reasonably.	A	4 (23.5%)	11 (64.7%)	2 (11.8%)	0 (0%)	0 (0%)	17 (100%)
	B	2 (10.5%)	10 (52.6%)	6 (31.6%)	1 (5.3%)	0 (0%)	19 (100%)
14. I have learned much from this class.	A	5 (29.4%)	8 (47.1%)	4 (23.5%)	0 (0%)	0 (0%)	17 (100%)
	B	1 (5.3%)	5 (26.3%)	9 (47.4%)	2 (10.5%)	2 (10.5%)	19 (100%)
15. The overall quality of this course is excellent.	A	5 (29.4%)	11 (64.7%)	1 (5.9%)	0 (0%)	0 (0%)	17 (100%)
	B	4 (21.1%)	8 (42.1%)	5 (26.3%)	2 (10.5%)	0 (0%)	19 (100%)
Total	A	61 (23.9%)	160 (62.7%)	34 (13.4%)	0 (0%)	0 (0%)	255 (100%)
	B	49 (17.2%)	116 (40.7%)	91 (31.9%)	22 (7.7%)	7 (2.5%)	285 (100%)

Table 10.1: A comparative evaluation form between L2-translation combined syllabus and traditional L2 syllabus

As can be seen from Table 10.1, many more students in Class A gave a highly favourable consideration to their course (100% on the agreeing side for the fifteen questions as a whole) than those of Class B to theirs (only 90% on the agreeing side). A word of caution is readily in hand: As the questionnaire did not specifically address the differences between the translation and non-translation paradigms, the actual differences observed in student evaluation may reflect other aspects of the syllabus and the teaching as well, plus the fact that translation was only partially used for Class

A, which further mitigates its effect. Nevertheless, since the two classes were taught by the same instructor and the students from the two classes were in fact of the same age and with approximately the same level of competency in English, it should not be too farfetched to say that the computational approach, and to a lesser degree the task of translation, may have contributed to the students' favourable evaluation.

Note that for Question 11, *The instructor is well-prepared for classes*, the number of students from Class A agreeing to this statement **very much** (29.4%) is nearly double to the number of those from Class B (15.8%). This could be due to the fact that, in a computational approach, the instructor often has to prepare online teaching documents very carefully and thoroughly, like that partially shown in Figure 10-2 (the original document was much longer, including examples), which gives the sense of well-preparedness on the part of the instructor. Also, based on Question 14, a lot more students from Class A considered they learned a lot from the course than those from Class B (76.5% vs. 31.6% for the first two agreeing categories). It can be surmised that the computational approach has enhanced students knowledge not only in terms of language, but also in terms of computer know-how, as well as skills in using concordancers and MT engines, etc. Hence the superior feeling of achievement for Class A in contrast to Class B. One area, however, where Class A gave less regard to the instructor than Class B did, was in terms of student-teacher interaction (Question 4), where only 5.9% of A agreed to the statement strongly; whereas 21.1% of B agreed. The reason may be that the instructor actually interacted more with Class B in the traditional classroom, and may not have had enough interaction with students in the less congruent computer lab environment.

It can be seen then that questionnaire investigations seem a viable way for evaluating the advantages and disadvantages of a computational, L2-and-translation combined syllabus, if the questionnaire is specifically and carefully designed for such a purpose. This section offered a brief example of student questionnaire analysis as a means for evaluating L2 teaching syllabuses. There are other conceivable ways of evaluation which can be further explored, for example, evaluation by testing, by computer-based user modeling techniques, by interviews, etc., which will not be furthered investigated in this thesis.

10.5 Summary and conclusion

This chapter approaches the problem of application of the proposed integrated model for translation and language teaching by first examining the issues of syllabus design on which the classroom implementation model must be based. We first looked at kinds of syllabuses for teaching language, notably those along the structural

specifications thread and those along the functional specifications thread. We then talked about translation teaching syllabuses by proposing a novel way of classifying, i.e. by distinguishing linguistic, cultural and technological syllabuses. Based on these two threads of discussion, we pondered the question of what a syllabus would look like which combines L2 teaching and the teaching of translation into the L2 in a computerised environment. It is suggested that a flexible kind of syllabus with multi-dimensional interactivity between the instructor, the students, the hardware and software is desirable for the implementation of such a model. Finally, an evaluation of a relevant course to date was also offered to give some idea of student response to such a course. After setting out the principles of syllabus design, the next chapter goes on to explore the more specific issues of classroom implementation.

Chapter 11

Implementation Issues

In this chapter we consider the implementation issues of the translation competence and syllabus design models we have been developing for students learning to translate into a second language. We will focus on the translation department of a small university in Taiwan – Chang Jung University (CJU).[62] The environment of the university and the students and the main course offered by the department are introduced first. A description of the hardware environment and software facilities directly relevant to the translation courses are then offered. A discussion then follows focusing on the actual conduct of the translation classes with special reference to the computer assisted instruction component.

11.1 Department and course

CJU is a small university in the south of Taiwan. The Department of Applied Languages for Interpretation and Translation (DALIT),[63] as of the end of 2001, comprises ten undergraduate classes with altogether over 500 students. The author, an associate professor with this Department at the time of writing, is one of the 15 full-time and 18 part-time instructors in languages and translation, etc.

The curricular structure of the main four-year course offered by the Department is briefly shown in Table 11-1, which contains a succinct analysis of the modules offered across the four undergraduate years (See Appendix D for full lists of names for modules).

[62] Sometimes the name Chang Jung Christian University is used officially as an English translation for the university's Chinese name – chang-rong da-xue, which does not include words referring to the university's nature of Christianity.

[63] The literal translation of the Department's Chinese name is simply Department of Translation, however.

Module group	Components	Credit hour percentage
Translation modules	General translation	12%
	Professional translation	
	Sight translation	
Interpreting modules	Consecutive interpreting	8%
	Simultaneous interpreting	
Language modules	English	55%
	Foreign Languages	
	Mandarin	
Content modules	Introduction to the computer	13%
	Western literature; Children's literature	
	ELT methodologies	
	Elementary medical science	
	Politics	
	International relationship	
	Management science	
	Philosophy of life	
	Broadcasting	
Common modules	Military training	12%
	Physical education	
	Campus service	
	Common knowledge modules	

Table 11-1 Analysis of course modules in the Department of Translation, CJU

The first conspicuous phenomenon about this curricular structure is the small proportion of translation and interpretation modules in the whole curriculum: There are only 20% of translation and interpreting modules among all the courses offered in this department of translation. A little more than half of the modules (55%) are devoted to language teaching, especially to the teaching of foreign languages, which include, apart from English, Japanese, German, and French. The remaining quarter of the course is divided between the content modules (i.e. those which teach the knowledge of a subject other than language or translation) and the common modules, which include PE, military training, and campus cleaning service, which are obligatory to Taiwanese college students regardless of departments.

The curriculum structure presented in Table 11-1 actually renders empirical support to the translation teaching model described earlier in Figure 3-2, where translation competence was shown to be built upon linguistic competence. It is quite

reasonable then, for the curriculum to devote half of the courses to the teaching of languages.[64] The other support this curriculum model lends to the competence model proposed in Figure 3-2 is the rich variety of content modules offered, which help build up socio-cultural competence on the one hand (i.e. contributing to the linguistic competence), and enable the trainee translator to deal with as many subjects of learning as possible when translating in professional settings (i.e. contributing to the translation competence; see Figure 3-1). Both kinds of competence are prescribed by the model explained in Figure 3-2.

Hung (1996) introduces two kinds of translation courses in Hong Kong. The traditional type of translation course is strongly literature-oriented for either language or translation modules. The professional translation model, on the other hand, has a course structure very similar to that presented in Table 11-1. According to Hung, the Higher Diploma in Translation and Interpretation offered by the Hong Kong Polytechnic has the following course structure:

- Year 1: Languages studies, General translation and interpreting, General knowledge
- Year 2: Languages studies, Politics & administration or commerce & industry or Mass media and interpreting, General knowledge
- Year 3: Arts & literature or Science & technology studies, Politics & administration or commerce & industry or Mass media and interpreting, General knowledge, Group project

which, on the whole, correspond to the four main groups of modules for DALIT, CJU analysed in Table 11-1 (i.e. Translation modules, Interpreting modules, Language modules and Content modules). It can be seen then, that the DALIT's main course is a modern and professionally oriented one, presumably more congenial to technologically enhanced educational features, such as computer assisted language learning or translation teaching.

11.2 Environmental issues

The section discusses the "hardware" component of the teaching environment of DALIT (here referring to the environments where teaching and learning activities take place). The venues for teaching currently used by DALIT include: traditional

[64] Though in fact a student in the Department does not have to take all these language modules, since, for instance, she only has to choose a foreign language to pursue among Japanese, German, and French.

classrooms, the computer lab, and asynchronous distance learning environment.[65]

11.2.1 Traditional classroom

Traditional classrooms are still the main (or only) venue for instruction for most teaching fellows at DALIT, though not for the present author. Other teachers of relevant modules (e.g. English listening comprehension) also make use of the audio-video labs, though, again, not the present author mainly because there are no computers in the audio-video lab.

For the current author, the traditional classroom is used when the class session involves explanation of textbooks or handouts and plenty of blackboard writing. When the class session involves exercises which need to be done with a computer, the session is moved to a computer lab.

11.2.2 Computer laboratory

DALIT, CJU is fortunate enough to have its own computer lab (i.e. not sharing with other university departments or students). The Department, however, does not bother to employ a full-time or part-time computer officer, probably because the lab is subsumed in the school's network and the Department does not have to worry about network connection or hardware and software maintenance, which is taken care of by the University's computing centre (when a serious problem arises). As there is no sitting-in computer officer, however, any Department staff member wishing to use the lab for teaching purposes is usually on his own devices and at his own risks regarding software deployment, use and maintenance. Consultation and help from the University computing centre is remote and far from being enthusiastic.

At the time of writing, the Department computer lab has only twenty-five new computers equipped with Intel Celeron 650MHz processors shipped with preinstalled Microsoft Windows 98 (Traditional Chinese) systems. Two or three students working at one computer is the norm since a translation or English writing class often consists of more than forty people. The Department is on the way to purchasing more computers, however, to double the number of computers in the lab, which is scheduled to happen in 2002.

Each computer in the lab is also installed with a "reborn card", which is a kind of PCI expansion card that interfaces with the hardware to provide more administrative

---

[65] There are also some audio-video labs for video viewing and listening comprehension, etc. in the Department. Since the author is the only Department staff involved in the current project, and the author never makes use of the audio-video labs, they are excluded from the discussion of this thesis.

capacities and security measures, such as avoiding the damage to the system or the loss of data. An advantage of the reborn card is that the settings of the computer system and application software will not be accidentally or intentionally changed by the user, since all these settings will be restored every time the computer is rebooted. A disadvantage, however, is that users cannot save data permanently in the hard drive, as they will be wiped out also upon the next reboot. The user wishing to save any data processed must prepare her own floppy disk, whose space is rather limited and the access speed is low. Though there is another option of saving data to the user's network space, most student users in the Department, however, being used to the PC Windows 98 system, do not know how to make use of their personal space in the University's network server.

All computers in the lab are individually connected to the Internet, as well as being controlled by the main computer, which only the instructor can access, through a local area network (LAN). The key software which provides the immediate teaching environment for the lab is a LAN broadcasting system called LanStar[66] (Figure 11-1). LanStar allows the instructor to take control of (or monitor) the screens of student computers to make presentations or conduct online conferencing (Figure 11-2), etc. It also allows file transfer between the instructor's and the students' computers (Figure 11-3). LAN provides an excellent environment for conducting CAI (computer assisted instruction) lessons, as it offers not only networked computers, but also human supervision at the same time. However, one weakness of LAN is that all people must be in the same place at the same time to make the system work. For learning materials and activities which can allow students to access at any time and from any place, distance learning is the solution, to which we now turn.

[66] Homepage at: http://www.goldensoft.com.tw/, which is a Chinese page.

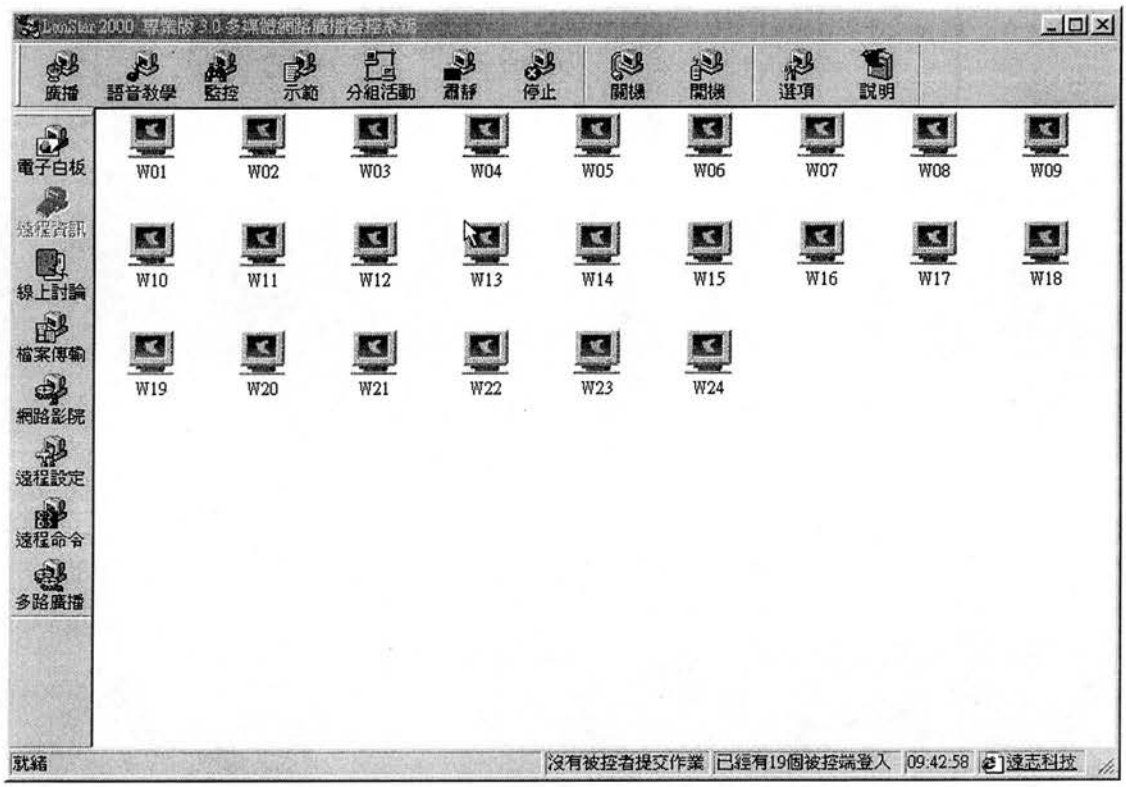


Figure 11-1 LanStar: The control panel

Figure 11-2 LanStar: The chat program

Figure 11-3 LanStar: File transfer

11.2.3 Distance learning

Although it is still not a common practice to deploy intelligent tutoring systems on the Web as a distance learning component, the possibility should be contemplated, for the simple fact that the ITS is best suited for autonomous study, and so is a distance learning system. The ITS conducted in an asynchronous distance learning environment should provide students with the best off-class opportunities to learn.

The computing centre of CJU has an in-house developed distance learning system, Asynchronous Web-based Instruction System (AWIS), which is available to use for its staff members. The author of this thesis works closely with the computing officer responsible for designing and maintaining this system to make it more useful for computer assisted instruction purposes. The author has been including distance learning as one of the components of his personal teaching schemes, applied to the teaching of translation, English composition, and other modules. The following sections illustrates the functions of the distance learning system through a Web-based English writing class, which is in many ways similar to a translation class in terms of management in a distance learning scheme: for instance, the assignment files being handed in for translation and composition are in similar formats, and the presentation, discussion and teacher feedback mechanisms are also largely the same. Furthermore,

since language learning occupies an even larger portion than translation in the curriculum of DALIT, as we saw in Chapter Ten, helping to build up the linguistic competence leading to translation abilities, the discussion on a Web-based English writing class is directly relevant to this thesis.

11.2.3.1 Introduction

The English writing class is an important part in the whole curriculum of EFL (English as a Foreign Language). A course of EFL composition following the traditional rhetoric approach usually centres around the skills in developing various patterns of paragraph (e.g. Nadell et al.1997), and a traditional English composition class session may be conducted in the following way:

1. Teacher explains skills to learn (e.g. *narration* or *description*)
2. Teacher assigns topic of the day
3. Students write composition
4. Teacher collects compositions and marks them at home.
5. Teacher returns students' compositions at next session with correction/feedback.

However, this approach was criticised as too 'product oriented' and the students do not get enough or the right kind of help from the teacher. An alternative approach was proposed known as process writing (See White & Arndt, 1991). The class procedures under the process writing approach may be:

1. Teacher provides stimulations to generate ideas
2. Students brainstorm or discuss based on materials provided by teacher
3. Students write drafts.
4. Teacher responds to drafts.
5. Students revise drafts before handing in.
6. Teacher marks compositions of class and returns them in next session.

The process writing approach emphasises the "process of discovery" (Raimes, 1983: 11) -- through peer discussion and interaction with the teacher, the students discover new ideas and use new language forms to express these ideas. A typical modern writing class will benefit from both traditional and recent methodologies in teaching writing and probably contains elements from both, for example, explanation of paragraph development skills, presentation of materials and discussion before writing, and teacher support during the actual writing.

With the support of modern computing and networking technology, the teaching of EFL writing has also gained new momentum and is able to offer new solutions to the previously challenging problems such as presentation of learning materials, access to reference materials, and assignment management. The English composition class conducted by the author is a regular on-campus course rather than a distance learning class. The class has physical gatherings both in a traditional classroom and in a computer laboratory, but the students access the distance learning pages, when required, not only during the class times in the computer lab, but also at home or in school outside of the class hours.

11.2.3.2 System overview

The AWIS has a log-on interface as shown in Figure 11-4, which is shared by both students and teachers, each logging on to their separate systems. This section mainly introduces the teacher's working environment, but the students' environment is essentially the same in terms of both interfaces and contents.

Figure 11-4: The logging-on interface of AWIS

After logging on to the teacher's personal space, the teacher can choose one of the

classes taught by him to administrate, as Figure 11-5 shows.

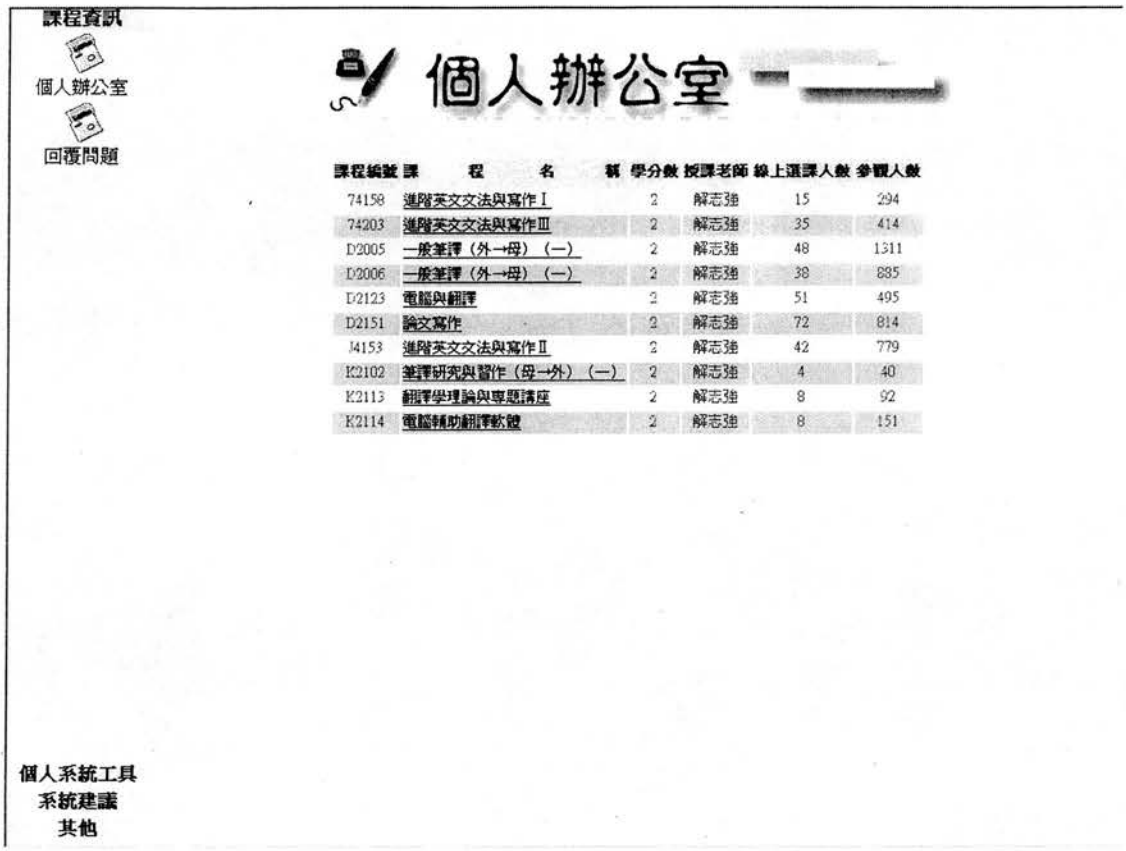


Figure 11-5: Display of available classes

Suppose we choose the first class, English Grammar and Composition. The interface then switches to a Web space belonging exclusively to this module, as Figure 11-6 shows.

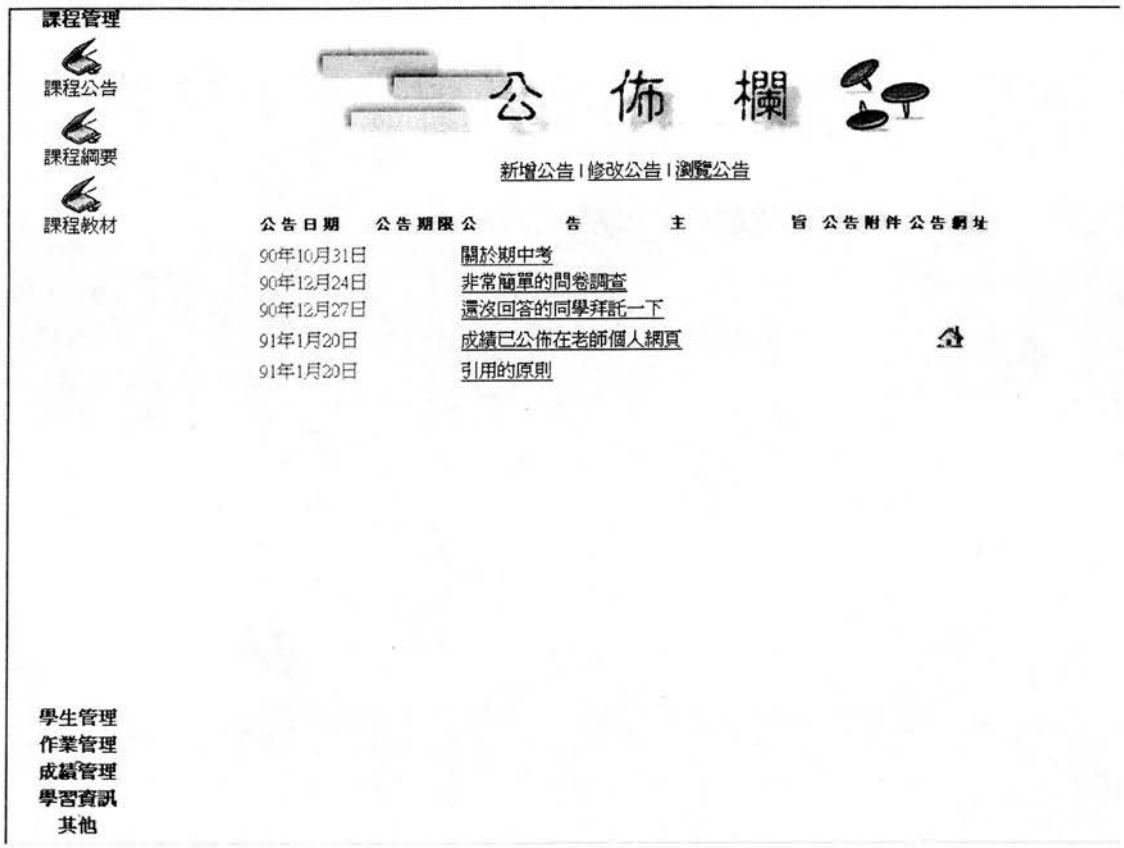


Figure 11-6: Individual class space

Note the line of six (green) buttons to the left of Figure 11-6 (one moved to the top upon clicking, five remaining at the bottom): These are the function buttons the teacher can use to administer the class. The pressing of each button expands the administration job into a number of sub-functions. For example, on the top-left corner of Figure 11-6, we see three things a teacher can do with respect to Course Administration. If we press the third button, the following interface comes up, which hosts the main teaching materials for each class session:

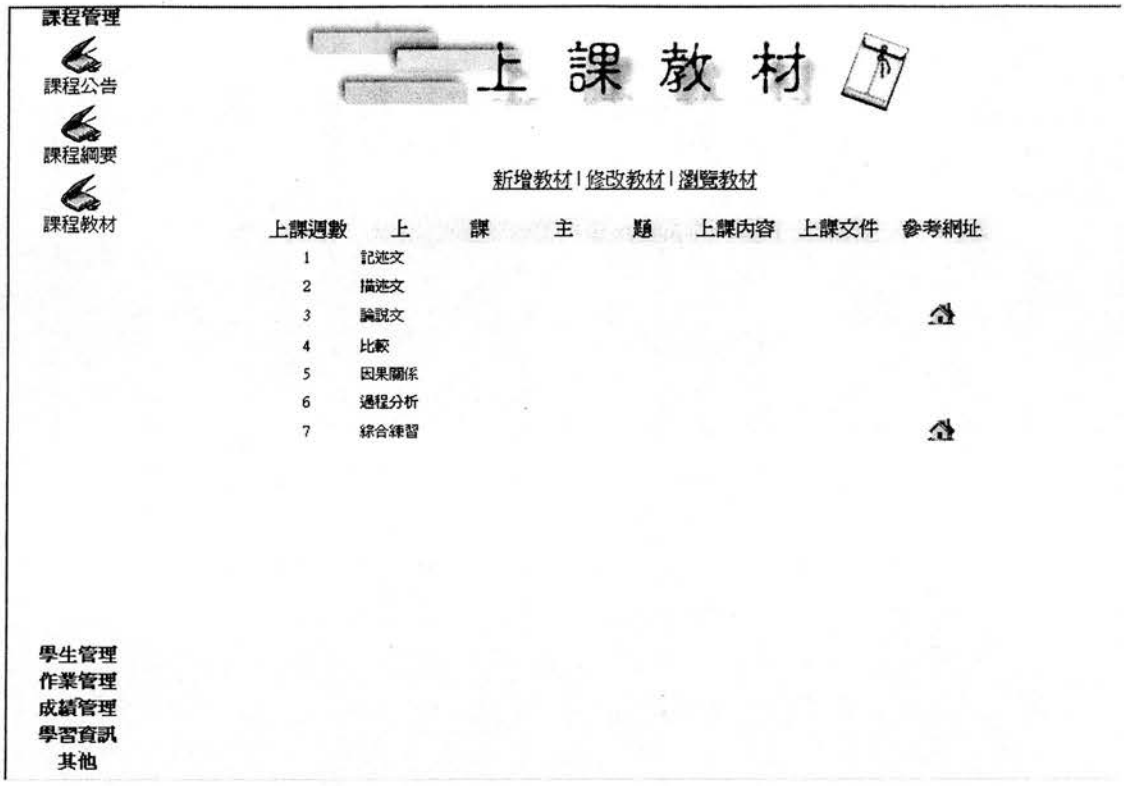


Figure 11-7: Teaching materials section

There are three kinds of teaching materials that can be included in each week's contents of teaching: 1. a summary of content to be learned for that week (This is in plain text format typed in a preset box); 2. supporting documents which can be in either Word or HTML form; 3. a referenced Web page -- an example is shown in Figure 11-8:

Figure 11-8: A referenced Web page

Note that if a Web page being referred to is outside of the campus network, then the speed for access can be slow depending on the kind of network installed, the location and complexity of the Web page, and the pedagogical task involved. Some Web pages can be saved in local disks and called out for use later in entirety, but some cannot, especially those involving interactivity, like remote database search, online translation, or video streaming. Slow speed in accessing data can cause students' frustration and upset a lesson plan.

The next important button is the Assignment Management button, which expands into three sub-functions: 1. giving out assignments; 2. correcting assignments online; 3. answering students' questions, if any. Figure 11-9 shows a writing assignment being given out by the teacher. Supporting materials can be attached either in Word or HTML form and Web pages can be referenced as well in the assignment-giving page.

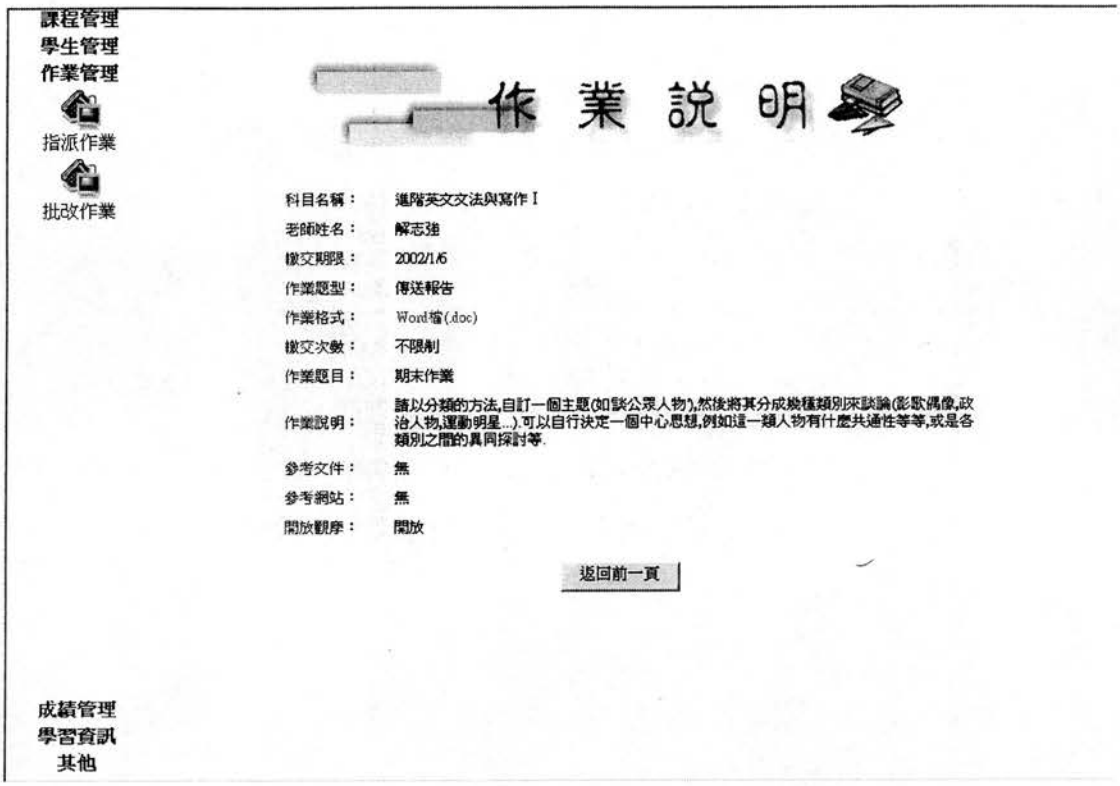


Figure 11-9: Giving out an assignment

After the student has logged on to her personal space, the student can see the assignment uploaded by the teacher, do the exercise and submit online in the form specified (currently the system supports Word files submitted online or plain text typed in preset boxes). The teacher then sees a screen like Figure 11-10 in his Assignment Marking interface.

Figure 11-10: Assignment submission list

The teacher clicks on one of the submitted assignments to open a document like the one in Figure 11-11, which happens to be a Word file (The other possibility is a plain text file). The current system only allows the teacher to give grades and comments in the columns provided by the system interface as shown in the upper half of Figure 11-11.

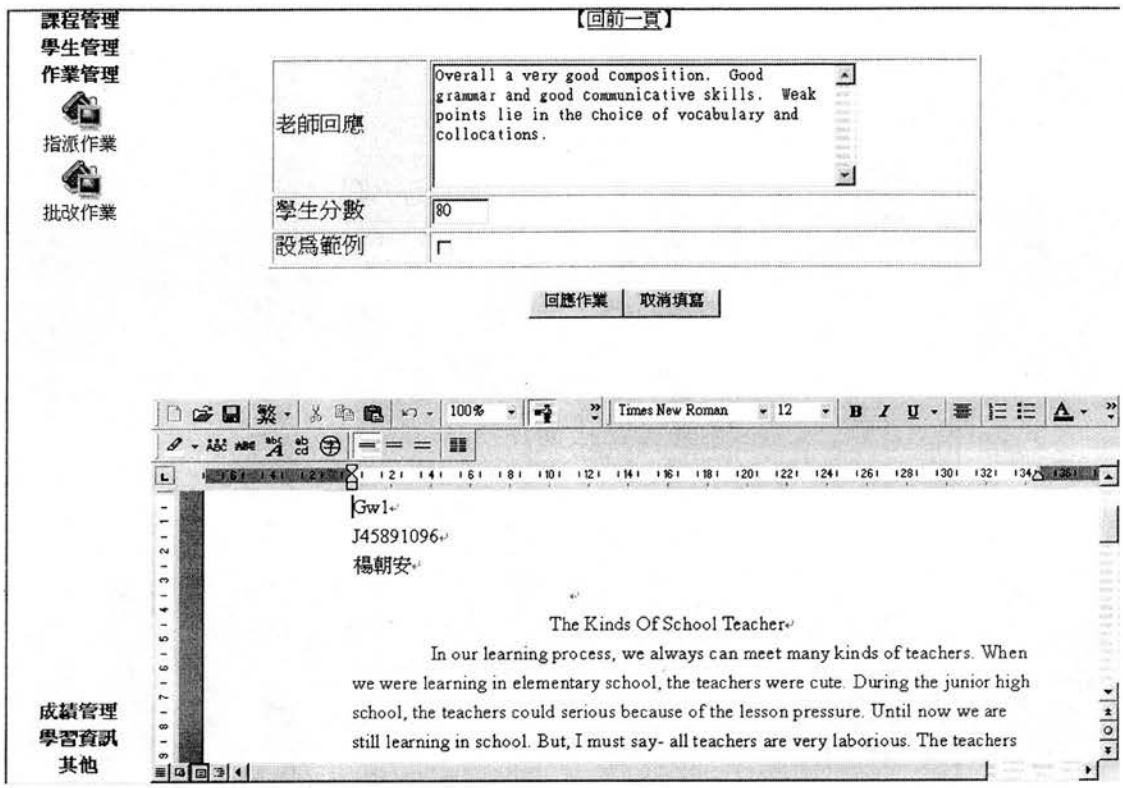


Figure 11-11: Online grading

The system is being upgraded, however, to allow the teacher to correct the assignment online as well. That is, in future the teacher will be able to place comments on the Word file itself, as Figure 11-12 shows, and save the file in the original space for the student to retrieve later and see the teacher's corrections and comments at relevant places in the original document.

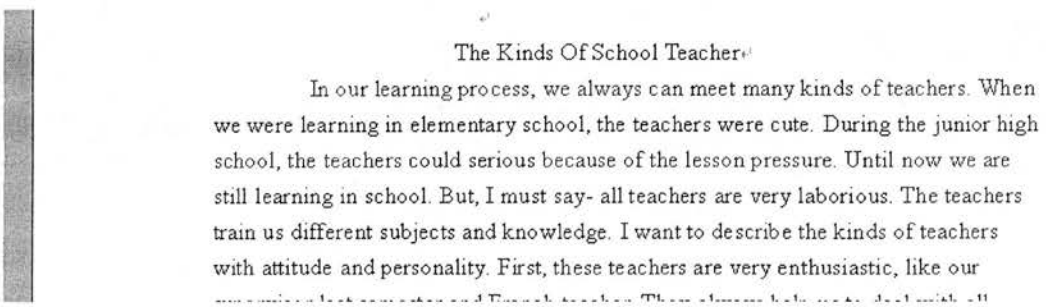


Figure 11-12: Online correction

A final important function of the system is in the Learning Information area, where the student's access record is shown and can be used to evaluate the popularity of the system and the student's degree of commitment, etc., as Figure 11-13 shows.

學生班級	學生學號	學生姓名	學生帳號	上課次數	上課時間(分)	發問次數	瀏覽次數	瀏覽時間(分)
翻譯系四甲	J45891096	楊朝安	marvin0711	8	237	0	1	0
翻譯系四A	J45891177	孔華	J45891177	14	155	0	1	0
翻譯系四	J45891389	陳佳慧	J45891389	22	585	0	2	0
翻譯系四	J45891402	陳姿君	Sylvie23	28	714	0	2	21
翻技四	J45891428	陳美君	meichun	33	693	0	7	1
翻譯系四A	J45891444	彭蘭珍	J45891444	14	513	0	12	62
翻譯系四甲	J45891452	黃媛筠	ann	10	152	0	1	0
翻譯系四A	J45891460	黃嘉類	j45891460	12	340	0	0	0
翻譯系四甲	J45891486	葉至芹	anna	14	326	0	1	0
翻譯系四甲	J45891494	劉綠琦	j45891494	26	607	0	0	0
翻譯系四A	J45891525	蔡菁菁	j45891525	11	410	0	1	0
翻譯系四	J45891533	蔡雅雯	J45891533	17	624	0	2	4
翻譯系四A	J45891567	鄧婉怡	J45891567	8	286	0	3	64
翻譯系四A	J45891591	賴若盈	j45891591	36	805	0	11	10
翻譯系四A	J45891606	謝佳雯	J45891606	8	218	0	0	0

Figure 11-13: Cumulative access information

There is also a discussion board in the Learning Information space, which the students (as well as the teacher) can use to discuss anything related or unrelated to learning.

11.2.3.3 Advantages

The advantage of a Web-based writing classroom as compared with a traditional class are:

1. Teaching materials presentation: In a traditional classroom, the teacher usually has textbooks, handouts, slides, and the blackboard at his disposal. The teacher usually has to borrow extra equipment if he wishes to use audio or video tapes in class for prewriting activities. But in a computer environment, all these media can be organised into a single Web page. Presentation of teaching materials can be more compact, organised, attractive and efficient.
2. Resource finding: A portable electronic dictionary which can be used in a traditional writing classroom is no comparison to the vast pool of resources offered by a powerful connected computer. Dictionaries, thesauri, collocation dictionaries, search engines, concordancers, online machine translation, library catalogues -- all these useful references and many others are at the students' disposal once they are sitting at a computer connected to the World-Wide Web.

And the correct and effective use of these facilities is best introduced and supported by the teacher in a computer lab or on a distance learning site.

3. Assignment management: Submitting assignments in electronic forms does away with the sorting and handling problems associated with paper work altogether, and also makes it easier for the teacher to store, manipulate and do research with students' writings. Submitting the assignment file through a Web page has further advantages over submitting by disks or email. Submitting by disks or email often causes confusion, uncertainty and worry, and corruption of files.[67] On the contrary, submitting through a Web page to a designated location in a Web server ensures safe delivery and the orderliness of the assignment files. The instant acknowledgement by the server upon reception of the file, on the other hand, reassures the student of the successfulness of the transaction.
4. Teacher feedback: By collecting the student assignments more speedily and in a more orderly fashion, which are typed and easily readable documents as compared with handwritten scribbles, the teacher is able to give better feedback and in a timely manner. Electronic files also make it possible for the teacher to do various kinds of task on the assignments using computer facilities or programming languages and produce useful statistics or generalisations.
5. Interactivity: Online discussion in the Web forum can encourage the most timid students to ask questions or express opinions due to its anonymous nature. An ever-existing discussion forum allows the students to talk to each other over the assignments or class materials throughout the week and clarify any doubts or misgivings.

11.2.3.4 Future improvements

There are a number of points that can be further improved in the current online system. For example, the system can include a campus Intranet RealServer (Chen, 1999) to deliver video clips linked to by the teacher's presentation pages in order to make the students' access speed faster. Currently any video clips referred to by the teacher's Web page are streamed through the original locations, wherever they are, and the access speed can be very slow. An inexpensive, or even free, basic video server, having cleared the copyright issues and stored the videos locally, can make the streaming much faster and the videos much more pleasurable to see.

Secondly, an integrated online chat program can also be developed for the

[67] In fact, before the AWIS system became available, the author used email as a means for submitting student assignment, and confusion and mistakes were a fact of life. The most upsetting thing is that students often submit repetitively or lose their emails or forget to attach files to emails.

system so that students and the teacher can discuss any topic-related or general issues online. Current online discussion is only possible when the class is physically conducted in the Department's computer lab, which is installed with the LanStar broadcasting system with integrated chat functions (Figure 11-2). In future the online chat should also be made possible directly through the distant learning Web site, where the student can chat with anyone (including the teacher) available online. A program like this can already be seen in the New School Online University Web pages,[68] for example, where an available staff member can chat with any visitors having questions.

A final area where we would like to see the AWIS improved is the assignment management module, especially the machinery for the provision of teacher feedback which is of crucial importance for writing or translation classes. As mentioned before, the online correction system would soon allow the teacher to put comments on the students' Word document itself (Figure 11-12). However, adding comments to Word documents (somewhat like adding in a footnote) is not a particularly efficient or pleasant thing for the instructor. A more dedicated marking system seems desirable if we want to reduce the teacher's labour and to make the marking results more meaningful and reusable. There is a handful of software especially designed for student assignment marking purposes, one of which is Markin,[69] the starting interface of which is shown in Figure 11-14.

[68] http://www.dialnsa.edu/

[69] http://www.cict.co.uk/software/markin/index.htm

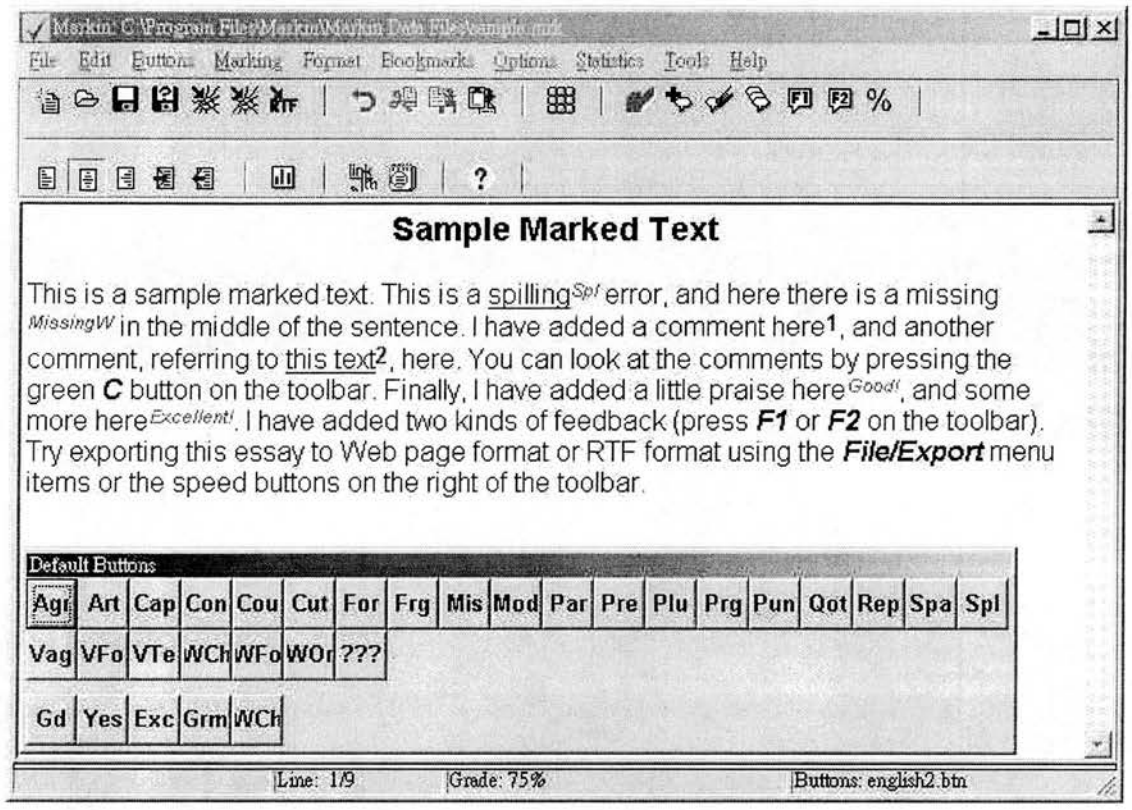


Figure 11-14: Markin -- a marking and annotation software

The reader will notice the rows of buttons at the bottom of Figure 11-14 (which are designed to be editable by the user). These are the tools for the teacher to easily insert error types and comments in the student text. For example, the button *Agr* stands for 'error in verb-noun agreement'. Once the teacher highlights a piece of text and clicks this button, the highlighted text will be underlined and change colour, and the label *S/VAgreement* in red will tag behind the text as superscript. What's more, the teacher can export the corrected document as an HTML file to be displayed in a Web browser, which, in addition to the annotated text, will output the grade given by the teacher and simple statistics of the errors types and teacher comments, as Figure 11-15 shows.

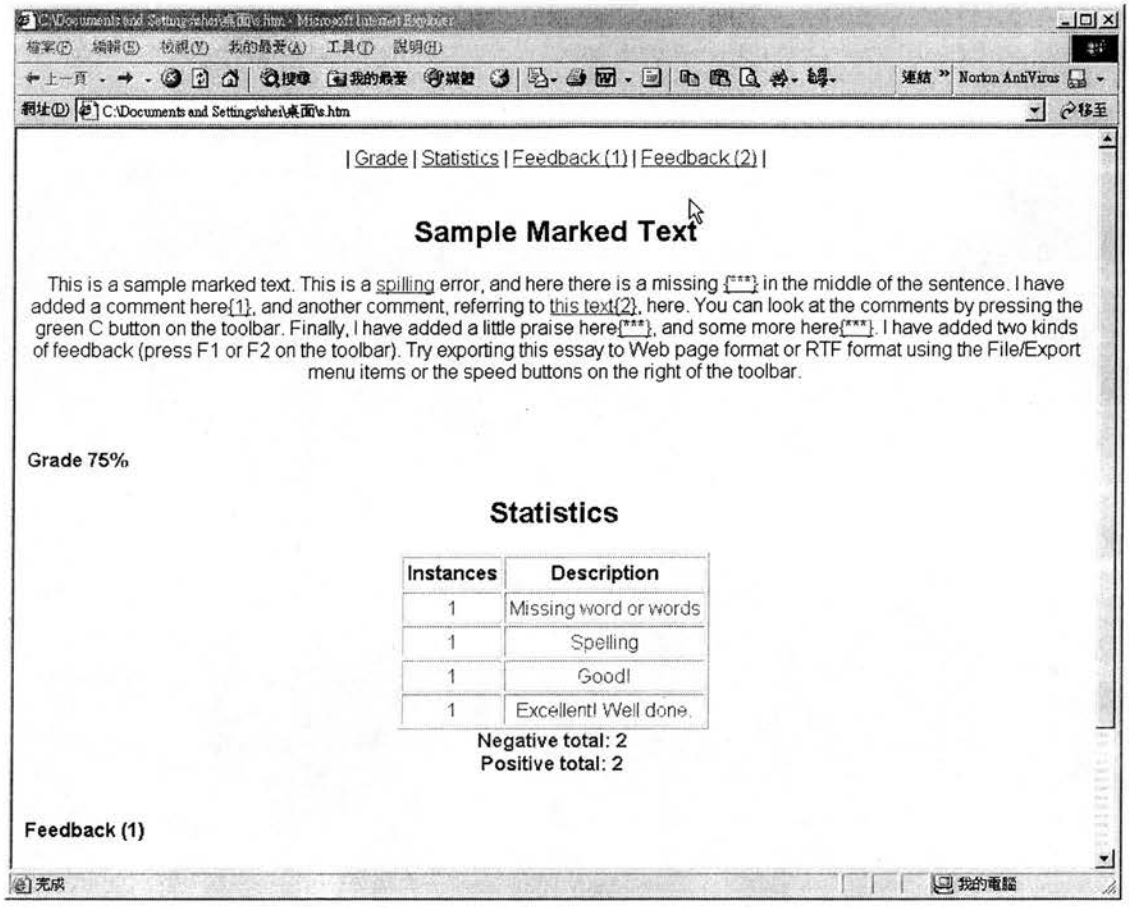


Figure 11-15: The HTML output of Markin

Unfortunately, it is inconvenient, in our present context, to download the student's assignment from the server, mark it with Markin in our local computer, and then upload the HTML file to the server again to be displayed on the Web. What we are looking for instead is a marking machinery which is integrated seamlessly in our Web-based system which has the same convenience as Markin's for adding in teacher comments. The statistics are also useful for individual students, but we are also aiming at more complex statistics for the entire class, displaying information such as common error types and the teacher's universal comments to these error types.[70]

With respect to peer emulations, the current AWIS system allows all students to see all other students' assignments and the names associated with them after the submission deadline. This has pedagogical values but infringes on student privacy. Future system should ask students, when submitting their assignments, whether they would like to make their pieces of work visible to others or not. In a nutshell, we are envisaging a marking system which can allow the teacher to easily put up well stocked Assignment Feedback Web pages incorporating richly annotated individual

[70] Reynolds (2001) offers a good discussion on similar functions for an online translation marking system.

works (some of which are publicly visible), and well-organised error analysis and comments, as well as statistics for the entire group.

As a final word for this section, it should be pointed out that instructors in translation are still a rare species in the field of education, compared to other disciplines such as language teaching. Many instructors in the Department are self-taught and self-trained translation teachers whose original specialties are in linguistics, literature or other related fields. Also, as pointed out in the beginning chapters of this thesis, the study of translation and the training of translators are strongly associated with the profession in practice. Thus translation teachers who are learned scholars as well as experienced field translators are extremely rare and should be keenly sought after by academic departments in translation. It is therefore important for a translation department to consider the distance learning mode which can allow a highly qualified translation teacher to be "shared" across institutions via the Web.

11.3 Software issues

This section discusses issues associated with the acquisition or development, installation, use and maintenance of software to be used in the Department's computer lab for teaching and learning purposes. The discussion covers all three types of software distinguished in Chapter Six that can be used to teach language and/or translation.

11.3.1 Development

Hubbard (1992) recognises the following components of procedure for developing CALL software:

- Activity type
- Presentational scheme
- Input judging
- Feedback
- Control options
- Help options
- Screen layout

The only way to come by Type B programs, like those discussed in Chapter Eight, seems to be for instructors to develop them in-house, since there is very little chance

of finding such software on the market which combines L2 learning with translation teaching. However, for an arts department like DALIT there are very few facilities and little human power for developing CAI programs. A look at Hubbard's list above would show how complicated the task is to develop full-fledged programs like the Translation Micro World discussed in Chapter Eight. The "input judging" component, for example, for the Translation Methods Tutor discussed in Chapter Eight, would involve very sophisticated computing to determine what translation method the student is most likely using when translating a given sentence. The other difficult component is the feedback session. For the Translation Micro World programs, if the feedback is to be intelligent enough to be of any use to the student, lots of algorithms must be written which monitor the student's progress, searches the corpus and makes intelligent selections in order to present the best illustrative material to the student in response to the infelicities in the student's writing.

By far the most difficult aspect of developing Translation Micro World applications, however, is the acquisition of language corpora, which play a key role in designing such programs. Large corpora like BNC and Bank of English are normally not to be sold to a Taiwanese institution in its complete form.[71] English-Chinese bilingual corpora are few and difficult to come by, and it takes huge time and efforts to build one's own corpora. Therefore, a few years' time is to be expected for the Department to set up a software development team consisting of translation teachers, linguists, computer experts and software engineers, though it is questionable whether the Department would have such plans and funding to do so or not. For these reasons, the immediately possible solution is to try out the concept of teaching proposed in this thesis using Types A and C programs, i.e. ready-made software which allows some space for authoring or modification to software usage of some sort.

11.3.2 Acquisition

Recall that the Type A and Type C programs which are closer to their respective extremities, are both purchased pieces of software. Type C programs are CALL software most of which should be developed by software companies employing language teaching specialists, along with other professionals, and thus should present little problem for use in an educational setting. Type A programs, on the other hand, include ready-made commercial translation-aids software such as the translation memory (TM) systems, which are not developed for educational purposes. The

[71] Users from a place like Taiwan can only subscribe to an online service to query the corpus database, not to get the entire corpus data file.

introduction of Type A programs into an educational institution prove to be problematic in some ways.

DALIT has purchased two kinds of translation memory systems: TRADOS and Transit. The purchasing of these pieces of commercial software is not only costly but also time-consuming. The educational institution must allow time for: placing orders, remitting money, delivery, and getting familiar with the software, evaluating it, designing courses based on it, including finding learning materials and teaching methodologies to go with the demonstration or practice of the software. In actual purchasing processes, especially for international trade, some unexpected delays or mistakes can happen. For example, some software companies ask for a certificate of "non-commercial use" from the university headmaster before it can grant an educational discount, which may take days to weeks to prepare. Also, the bank remittance could go to the wrong account, and the goods which finally arrive may be incomplete in some way and need negotiation and re-delivery. These all happened to us in DALIT when purchasing the above-mentioned software.

11.3.3 Installation and maintenance

Many kinds of commercial software nowadays include anti-piracy features which can make the installation and use of such software very inconvenient. For example, both TRADOS and Transit are shipped with a kind of anti-piracy hardware key called "dongles", each dongle representing a licence purchased. The dongle has to be installed and found when the application software is executed for the full features of the software to be released. When the dongle is absent, usually only demo features are available. The dongle itself is to be manually attached to the printer port, without any protection. There is absolutely no guarantee, therefore, that the dongles will always remain in place – they might be stolen by users of the computers or intruders to the computer lab, for example. Somewhat unreasonably, if a dongle is missing, one copy of licence for the software also becomes void, as the software company will not replace the missing dongle.[72] In DALIT, this serious problem was found very early on. At first all dongles were properly installed for students to enjoy the full features of the TM software. After the mysterious disappearance of a dongle, however, all dongles were recalled and only demo features can be accessed after that. This problem can only be overcome by the software companies adopting other anti-piracy tools, such as that in the form of a controlling card installed inside the computer, to prevent pilferage.

---

[72] If the dongle is damaged instead of missing, however, it should be replaceable with a charge.

11.3.4 Use

When using CALL or TWS (translator workstation) related software, the most unpleasant thing that could happen is for the application to crash in the middle of work. The crashing incidence seems to occur more frequently when the operation involves manipulation of richly stocked HTML pages, for instance, when translating with a translation memory application. Worse still, the crash of the application often leads to the crash of the system, which results in two further inconveniences – the sudden disappearance of all the opened application interfaces on screen, and the loss of data having already been processed.

In the author's experience of using one kind of TM software, the application hosting the translation interface crashes from time to time. Students have to press Ctrl+Alt+Del on the Windows 98 system to terminate the application, losing at least part of the data that have been processed. What's worse, if students accidentally press Ctrl+Alt+Del more than once, they are likely to lose all data and applications in progress since the system will restart and the reborn card will wipe everything out from the last session.

There are also other forms of inconvenience involved in using other application software such as that of CALL. For example, the Department purchased a number of copies of the "TeLL me More" multimedia English learning software. The software can only be used when the purchased CD is placed in the CD-ROM drive. This also creates some degree of inconvenience, since the retrieval, handing-out and recollecting of the disks during the class time are time-consuming tasks and error-prone.

11.4 Implementation of lessons

While most instructors in DALIT, at the time of writing, use traditional classrooms and audio-video labs as the main environment for teaching, the current author makes use of traditional classrooms, the computer lab, and distance learning facilities. This section discusses the issues associated with conducting a class based on the combined approach to L2 and translation teaching, which takes advantages of all the teaching environments discussed above. A module named General Translation which the author teacher, which runs two hours weekly, is used to demonstrate the design and implementation.

11.4.1 Lesson planning

In planning a language lesson, Purgason (1991) synthesises from previous researches and proposes the following considerations:

- What is taught is defined by student needs.
- What is taught is defined by real language use.
- Sound principles of learning are followed.
- Lessons are structured for maximum learning.
- The classroom atmosphere and interaction are positive.
- Learning is student-centered.
- Activities reflect actual communication.
- Activities balance accuracy and fluency.
- Activities encourage interaction, both between learners and texts and among learners.

The current author suggests that the following points for consideration in planning lessons be added for a translation module incorporating language learning, while recognising all the points raised by Purgason above:

- The materials used for translation should be able to reflect not only translation techniques but also awareness of idiomatic L2 usage.
- The learning activities should allow both the discovery of principles of translation and exploration in the idiomatic aspects of L2.
- The skills being taught to students should include not only those related to the finding of high-quality, communicative translation for a given source text, but also those concerning the effective use of electronic tools in locating idiomatic target text.

In discussing the structure of a language lesson, Richards and Lockhart (1994) focus on four dimensions (p. 114):

- Opening: How a lesson begins.
- Sequencing: How a lesson is divided into segments and how the segments relate to each other.
- Pacing: How a sense of movement is achieved within a lesson.
- Closure: How a lesson is brought to an end.

Purgason (1991), on the other hand, suggests that a lesson consists of a beginning, a middle, and an end; where the typical sequence for the middle is: presentation, practice, communication. The "sequencing" and "pacing" components of Richards and Lockhart seem to refer to the principle of structuring, rather than the actual division of a lesson into segments. Here, based on Purgason's suggestion and the implementation model proposed in Chapter Ten (Figure 10-2), we divide a lesson plan into the following components: presentation, practicing, discussion, and reorientation.

11.4.2 Presentation

As previously mentioned, the current author uses three channels for presentation: blackboard writing or slides showing in a traditional classroom, slide show and computer demonstration in the computer lab, and teaching documents downloaded through a distance learning Web page.

A unit of presentation in the author's General Translation class is, for example, about the skills of translating verb-noun collocations from Chinese into English. The presentation consists of slide show in the computer lab with the following outline:

1. Introduction to collocation
2. Categories of English collocation
3. Difficulties presented by collocation in translating
4. Analysis of student translation
5. Procedure for dealing with collocation in translation into L2
6. Introduction to tools for looking up English collocations

The same material is also placed in the relevant section of the distance learning Web site, the Asynchronous Web-based Instruction System in CJU, so that students can access it at any place and any time.

11.4.3 Practicing

Immediately after the presentation of teaching material, the practice session begins, which, following the same example used in the last section, consists of student using computer to analyse Chinese collocations in the source text provided by the instructor, and finding acceptable corresponding English collocations using the electronic tools provided or referred to by the instructor.

Two of the most useful tools for translation students learning English as L2 are

Web concordancers drawing from corpora and Web search engines backed up by huge collections of documents from the Web. This is well-document in Lindquist (1999), who discusses three kinds of resources to help the translator decide on acceptable English expressions:

- Newspaper corpora
- The CobuildDirect Corpus
- The WorldWideWeb

One important tool which Lindquist has not mentioned is the "collocation sampler" on the CobuildDirect Web page, which provides a list of collocates to the query word in order of statistical significance represented by t-score.[73] Thus for example, in the English collocation session, this author asks the students to read a short Chinese story and pick out ten verb-noun collocations and find their equivalents in English. So for example, for the Chinese collocation *he-tang* (literally 'drink soup'), the students are told to look up the Cobuild Collocation Sampler, typing in *soup* as the query word, where they will find the screen output like that in Figure 11-16:

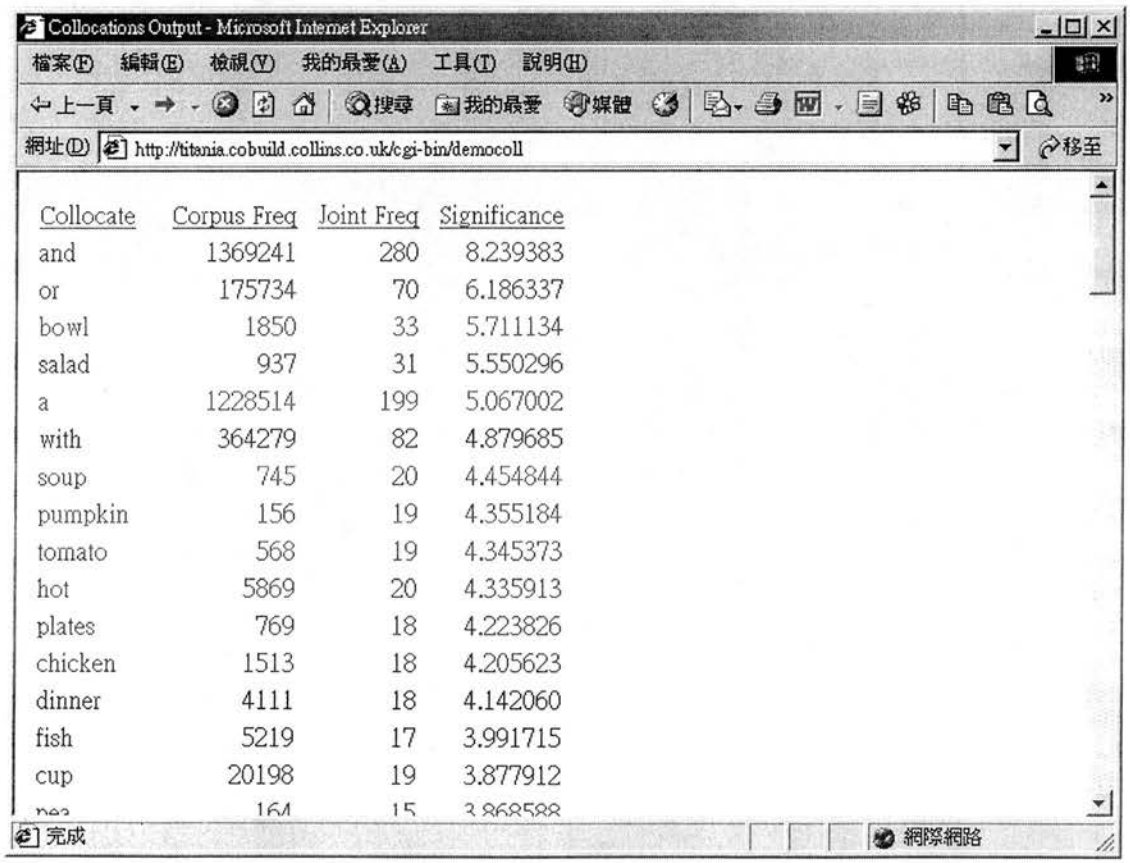


Figure 11-16: Cobuild Collocation Sampler partial output to query "soup"

[73] The function of t-score in this context is similar to the z-score discussed in Chapter Four.

They are then told to scrutinise all the output lines and locate any likely verb encoding the meaning of "putting things into one's mouth". While no incidence of *drink* or *have* or *take* can be found, *eat* is found lower down in the range of collocates provided, with the following statistics:

Collocate	Corpus Freq	Joint Freq	Significance
eat	3861	5	2.056835

Students are then told to use concordancers to find sentences containing the collocation *eat soup* in corpora and observe its behaviour in context, to see if it is a proper translation for the Chinese equivalent in the source text, and so on and so forth.

11.4.4 Discussion

After the lab session, the General Translation class will gather in the traditional classroom in the next class period, split into groups and discuss aspects of learning of the previous session, before advancing to the next unit. The points for discussion include:

- Translation skills learned
- Knowledge of language acquired
- Use of computer programs

There is also a permanent discussion board on the distance learning Web site, where the students can exchange opinions and make suggestions regarding the above points. There are usually more questions or doubts raised about the use of software and hardware than those in other areas. For example, for the Collocation Sampler illustrated in Figure 11-16, some students are confused about the positions of the collocates in relation to the query word in the corpus, some are curious about how the significance scores are calculated. Thus following the discussion, the instructor can arrange for some information on the extraction of collocation to be posted on the distance learning page as supplementary readings.

11.4.5 Reorientation

The last stage in the cycle of classroom implementation is the instructor's reorientation stage, where he reflects on the previous classes from the perspective of

classroom research (van Lier 1988, Richards and Lockhart 1994) on the basis of the following data:

- Informal observation conducted in class and during practicals
- Records of types and numbers of help students ask for while using computer software
- Student opinions and suggestions from the discussion sessions
- Any student modelling data gathered by intelligent tutoring systems used
- Student access records on distance learning Web site
- Messages students post on distance learning forums

The aim of the instructor is to analyse these data, make generalisations and formulate or modify design criteria (Hubbard 1992: 49) of software based on these generalisations. Here Richards and Lockhart's (1994: 12) definition of "action research" is relevant:

> Action research is used in this book to refer to teacher-initiated classroom investigation which seeks to increase the teacher's understanding of classroom teaching and learning, and to bring about change in classroom practice.

As argued before, the adoption of CAI programs in a curriculum introduces even more unpredictable variables to the implementation of a lesson – notably the crashing of systems to upset the lesson plan – and classroom research becomes even more sorely needed in order to identify and categorise these variables. Moreover, the definition of "classroom" has also changed with the advent of distance learning. As described above, the contemporary translation and language learning class can have at least three classroom settings – the traditional classroom, the computer lab, and the distance learning space. Classroom research should deal with data collected from all these settings, and the results generalised and the orientations made should also benefit all three venues for learning.

11.5 A lesson

As one of the first steps to realise a combined approach to teaching L2 and translating into L2 in a computerised environment, a class session in the General Translation module instructed by the author was devoted to pre-editing MT input. The direction of translation was from Chinese to English. The MT systems used was the Babel Fish offered by SYSTRAN at http://babelfish.altavista.com/translate.dyn. Here we will use a simple paragraph to illustrate the pre-editing procedure taught in this class. In (1) we have a randomly chosen Chinese paragraph from a randomly picked Web

page:

(1)
(a) Source text

十四歲許姓少女將歌手孫燕姿當偶像，認爲瘦就是美，控制飲食並運動減重，從四十五公斤減到廿八公斤，外型變成皮包骨，有天體力不支昏倒，被送到醫院急救，回到家仍繼續減重。

(b) Target text in English

14 year old of little surname few heroic women singer grandson swallow posture works as the idol, thinks thin is beautiful, the control diet and the movement reduces weight, reduces from 45 kilograms to 28 kilograms, the outlook turns the skin and bones, has day physical strength does not faint, is delivered the hospital first aid, returned proficiently still continues to reduce weight.

As we can see from (1b), the English translation by Babel Fish based on (1a) has quite a few problems caused by literal translation and mistaken analysis of structure. One of the most conspicuous problems may be the run-on sentence phenomenon and the subject-less sentences which are a peculiar characteristic of Chinese clauses. To rectify this problem, we need to amend the punctuation (i.e. change a Chinese comma to a period where we think an English sentence should be rounded up) and add some subject nouns to (1a). Thus we have the new set of input and output in (2):

(2)
(a) Source text

十四歲許姓少女把歌手孫燕姿當偶像，她認爲瘦就是美，少女控制飲食並運動減重，她的體重從四十五公斤減到廿八公斤，她的外型變成皮包骨，少女有一天體力不足昏倒，被送到醫院急救，她回家後仍繼續減重。

(b) Target text in English

14 year old of little surname young girl the singer grandson swallow posture works as the idol, she think thin is America, the young girl controls the diet and the movement reduces weight, her body weight reduces from 45 kilograms to 28 kilograms, her outlook turns the skin and bones, the young girl one day physical strength insufficient faints, is delivered the hospital first aid, after she went home still continue to reduce weight.

If we compare (1b) and (2b), we find (2b) to be generally better in that each sentence is clearly delimited and most are equipped with a proper subject. Some small adjustments to (1b) also produced improved results in (2b). For example, the erroneous 'few heroic women' (*shao nv jiang*) due to a wrong detection of

word-boundary was replaced by the correct 'young girl' etc. when the original character 將 (*jiang*) was changed to its synonym 把 (*ba*)[74].

Thus the students went through the process of modifying the input, observing the output, noting down the limitations of MT, thinking about how to modify the input further, and repeating the sequence over and over, until they obtained satisfactory results (or until they felt the MT output can no longer be improved).

The students were asked to hand in a list of MT flaws they observed afterward, which can be roughly classified into the following categories:

a. Structural limitations

- MT often deals with long sentences by (erroneous) chunking, translating each chunk first and then combining chunks to form an illogical sentence.

- MT deals with each sentence independently, i.e. cannot relate consecutive sentences in a meaningful way.

- Complicated noun phrases are problematic for MT systems.

- Most sentences produced by MT are incomplete, literally translated and prone to grammatical errors.

- MT often wrongly assigns the first word in the Chinese sentence as the subject of the target sentence.

- The longer the sentence, the more words are omitted, the more unusual or ambiguous words are used, the poorer the results.

- If the conjunctions in Chinese are omitted, the MT system will not detect the relationship between the connected elements.

- MT frequently fails to detect the relationship between the modifier and the modified, and the relative positions of these two often have to be adjusted to achieve better results.

- Chinese sentences consist of "a lot of verbs", which are difficult for MT systems. For example, MT cannot correctly detect the boundaries between consecutive verbs.

- A Chinese clause often consists of two (short) English sentences, but MT fails to segment these types of clauses and translates them into single peculiar sentences.

b. Lexico-semantic limitations

- The MT system used is helpless in determining verb tenses, the number of nouns and parts of speech, references of demonstratives, etc.

---

[74] For some unknown reason, however, the originally better 'thin is beautiful' was replaced by the worse 'thin is America' (though these are represented by the same characters in Chinese) even though no adjustment has been made in this respect.

- MT systems do not really "understand" the meaning of the Chinese sentences, which can be seen from the fact that literal (and often anomalous) translations prevail.

- MT cannot decide correctly which Chinese character is to be translated into an English preposition.

- Many function words in Chinese have more than one meanings, but MT mostly uses only one (most frequently used) meaning for each of these words in all situations.

- Some words which have meanings are mistakenly translated with non-words with similar phonetic representations.

c. Idiomatical limitations

- MT systems cannot handle most Chinese idioms, metaphors, colloquial expressions, borrowings from other Chinese dialects, or older Chinese variants.

d. Cultural limitations

- MT needs to improve its ability to select appropriate words for a specific background.

- MT does not consider "habit of speaking" and is not "humanised" enough.

e. Operational limitations

- MT is inefficient, as the paragraph given can be well translated by human within fifteen minutes, but it takes four to five hours to make an MT system produce a marginally-acceptable result through repetitive pre-editing.

- Sometimes the Chinese source sentence must be changed to a "distorted" style in order to get a better English translation.

In addition, students also provided some strategies for improving the performance of MT (in terms of Chinese-to-English translation) based on their pre-editing experience. The following types of pre-editing strategies can be identified:

a. Reorganisation

- In order to get better English translations, long Chinese sentences are best split into short ones, each having its own subject and verb.

- Changing Chinese sentences to "English style" sentences may achieve better translation results.

- As in human translation, we must chunk the Chinese sentence properly in order for the MT to achieve better results.

- Word order in source sentences must be adjusted to be like that in the target language.

b. Simplification

- Try to simplify Chinese sentence input, by using fewer and easier words

- The Chinese sentences to be translated should be clear and easy to understand, and without any omissions.

c. Addition

- Like subjects, the omitted objects of verbs in Chinese must be restored in order to generate complete sentences in English translations.

- Some addition of words to make existing words clearer is helpful.

- Time adverbs implicit in the Chinese text often has to be explicit added in order to get the desired counterpart in English translation.

- Where Chinese does not need an article, an article must nevertheless be added to nouns in the source text in order for the English translation to be correct.

d. Replacement

- Using pronouns in Chinese achieves better results than using proper nouns.

- Changing part of speech of a word can sometimes yield a better result.

- Replacing negatives with positives can gain better result, for example, changing *can not be defeated* to *can win*.

- Replace Chinese idioms with straightforward expressions.

e. Pre-translation

- When translating Chinese proper names, it is safer to manually translate them into English first before feeding them into the MT system; otherwise they may cause a confusion in meaning analysis[75].

- Before submitting the source text, imagine what the translation will be like first and make necessary alterations.

f. Punctuation

- Proper manipulation of punctuation can help MT achieve correct structural

[75] Chinese proper nouns consist of meaningful characters, and they are not distinguished from other kinds of words by mechanical means such as capitalisation.

analyses.

Some advantages for trainee translators doing pre-editing exercises are immediately obvious:

a. Technological awareness

Trainee translators get to know more about the nature of MT and become better prepared for collaborating with MT systems in the future.

b. Translation competence

From the processes of editing source text, translation learners are in effect reviewing and applying their translational skills while they are rearranging the words or structures of the source text etc to see how these might affect the MT outcome.

c. Linguistic competence

While scrutinising MT output in L2, trainee translators as L2 learners are actually reflecting on their own knowledge in the target language, possibly looking up dictionaries etc to verify their hypothesis.

d. Affective boost

Pre-editing the MT input is an engaging activity, as this involves the operation of some magical machinery, so to speak, especially to students from an arts background. Students are at once fascinated by the automatic translation machine and intrigued by the frequently anomalous output.

This class session of embedding L2 learning in a translation (into L2) task using an updated language engineering technology (MT), seems to offer a good chance for us to see the model presented in Figure 3.1 in a practical perspective, of seeing the practise of teaching translation into L2 with a view to enhancing L2 competence at work, which, in this case, is made possible by the interactive MT system offering instant feedback to learners' tests of their interlanguage hypotheses.

11.6 Summary

In this chapter, I introduced the main course offered by the translation department where the author works at the time of writing. It was observed that language modules included in this course occupy more than half of the credit hours in total, most of which are foreign language modules. This seems to verify the principal hypothesis of this thesis, which says that translation teaching must go hand in hand with language learning.

The teaching environment of the Department was then introduced, which includes the traditional classroom, the computer lab, and the distance learning

environment, with special focus on the distance learning setting, which seems a promising place to host integrated translation and language learning opportunities in the future, being able to offer unrestricted access in terms of time and space, and sharing of resources in respect of tools, materials, and instructors.

The chapter then went on to introduce the software environment for the computer lab of the Department, regarding the development and the acquisition of software, the installation and use of software in the computer lab, and the maintenance of it. A notable problem is the use of some piracy-protected software the Department purchased which causes a problem in the computer lab. The frequent crashes of Windows 98 when students use sophisticated translation software like translation memory systems also constitute a problem.

The chapter described the planning and implementation of lessons which integrate translation and language learning in a synthesised environment consisting of the traditional classroom, the computer lab, and the distance learning space. Lessons are a series of recycling procedures including presentation, practice, discussion and reorientation. Classroom research is essential for this kind of approach and learner data gathered from the computer lab and the distance learning site are valuable sources for pedagogical analysis, software adjustment and modification to lesson plans.

Finally, the procedure and results of a class session on MT pre-editing was reported. Although the learning task was overtly concerned with translation into L2, the learner was actually invited to put their L2 competence to test via a series of hypothesising, modification, observation and discovery procedures. The example session illustrates the nature of a learning task within the proposed pedagogical framework and the desirable learning environment.

Chapter 12

Conclusion

This chapter attempts to put together the theories discussed and models constructed so far, and to offer a synthesised view on teaching translation into the second language. As the reader may have noticed, this thesis has not focused on the explication of a particular computer-assisted learning system in great detail, but instead has emphasised literature review of related disciplines and presents relevant computational systems whenever possible to depict the perspective of teaching translation in the current and future context of information technology, focusing especially on the learner's acquisition of the idiomatic aspect of a second language while learning to translate into it. Because language technology and educational technology are expanding fast in all directions, and because the research in computer-assisted translation teaching systems is still in its infancy, the author believes this is a good way to contribute to the future of translation teaching and, to some extent, second language teaching. In this final chapter, the author will summarise the findings and speculate on the implications of these findings to the fields of translation teaching and L2 teaching, as well as pointing to some directions for successive research.

The following is a list of summarised points from the discussions presented in the previous chapters:

1. Translation into a non-native language is not traditionally a felicitous working mode but is a necessary practice for modern society.
2. Translation competence is built on linguistic competence with the addition of content knowledge, technological know-how, awareness of cultural differences, self-updating abilities, etc.
3. After the decline of Grammar-Translation method, translation is not used significantly to support second language learning, if at all. In fact, translation and L2 learning are mutually beneficial and more innovative ways of teaching L2 with translation in an computerised environment are conceivable.
4. Learner corpus analysis is a useful means for uncovering learners' weaknesses in L2 interlanguage, in terms of grammar and idiomatic usage.
5. The idiomatic-principle, i.e. using prefabricated items to construct sentences, is the first language production mode humans adopt. However, L2 learners

usually have poor knowledge of collocations and conventionalised forms of the target language.

6. Translation exercises can be used to draw L2 learners' attention to the idiomatic aspects of the target language, helping the learner to acquire more idiomatic usages of the language, which in turn helps to refine the quality of the learner's translation.
7. Translation teaching and second language learning can converge in the same corpus-based computer assisted instruction environment.
8. There is in theory a continuum of translation and language learning software which can be adopted by translation or language instructors, ranging from pedagogical use of translation memory software and MT systems to CALL software with a flavour of translation.
9. Type A programs are machine translation or machine-aided translation software which can be used to help students acquire professional skills in translation and, to a lesser extent, learn the second language being translated into, if especially arranged by the instructor in terms of materials preparation and implementation procedures.
10. Type B programs are specially designed programs which profess to teach translation but embed language learning facilities in the translation activities. These are called Translation Micro Worlds in this thesis, which make heavy use of corpora to build a world of bilingual text where the student can explore and discover skills in translation and idiomatic usages of L2, among other things.
11. Type C programs are by nature CALL programs whose intention is to teach language. They, however, make use of translation in some ways to achieve the goals for language learning.
12. A flexible kind of syllabus is suitable for such a computational approach to translation teaching. In such a syllabus, the goals of teaching guide the formulation of classroom and lab sessions, which are revisable by student feedback and teacher retrospection.
13. A combination of classroom instruction, computer lab practicals, and distance learning modules is a favourable environment for implementing the computational approach to teaching translation.
14. Classroom research based on the data collected from software monitoring, student-teacher conferencing and Web discussion forums is essential for ensuring the usability of the software, the appropriateness of the materials, the good coordination between the different learning environments, etc.

An overall model is proposed in Figure 12-1 to represent the above ideas graphically,

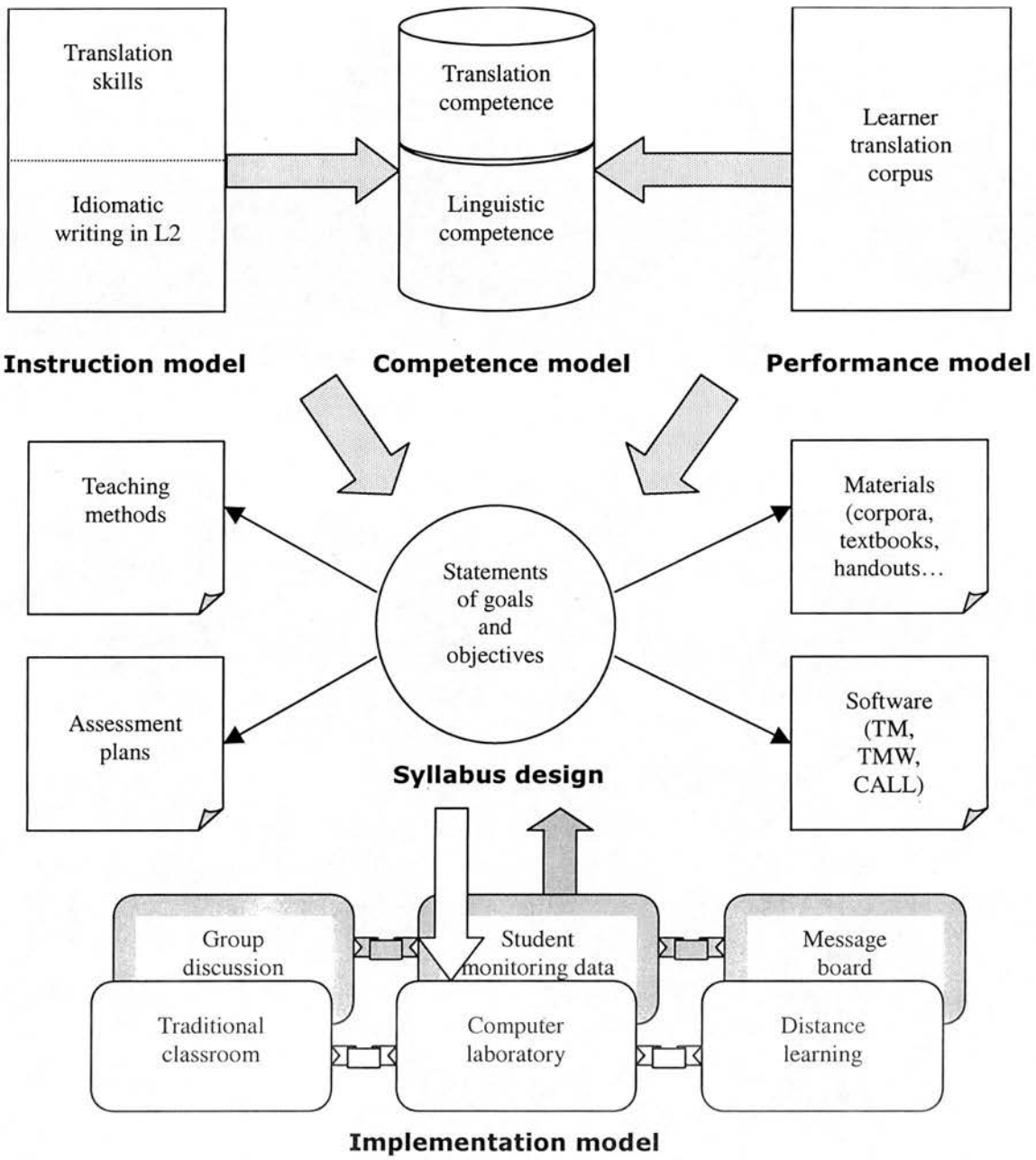


Figure 12-1: The overall model for teaching translation into a second language

The three models on the top of Figure 12-1 are explained below:

- The competence model: This is the ultimate goal for the trainee translator to achieve. Learners with excellent competence at this level can translate freely into the second language grammatically and idiomatically.
- The performance model: This is the model of learners' performance in relation to the competence model. From analyses of learners' translation corpus (performance), learners' abilities in using idiomatic target language (competence)

is approximated.

- The instruction model: Based on a comparison between the competence model and the performance model, the instruction model is proposed aiming to improve learner's abilities to translate into correct and idiomatic second language.

The next stage on Figure 12-1 is the syllabus design level. The three components at the first level act as input to the syllabus design level, where the performance model gives rise to an understanding of learner needs, the competence model represents the instructor's concepts about teaching, while the instruction model subsumes a selection of relevant teaching methodologies.

The syllabus design level itself includes the statement of goals and the breaking up of goals into smaller achievable objectives (Graves, 2000: 77), and the consideration of materials, software, teaching methods and assessment plans to be used in the course. The syllabus, when completed, is handed down to the classroom implementation level, the last component in Figure 12-1.

The classroom implementation model includes three kinds of environment where teaching and learning can take place: the traditional classroom, the computer lab, and the distance learning Web site. All the teaching and learning materials and activities conceived at the syllabus design level are allocated to these venues as appropriate to their individual settings. An important output from this level is the "feedback stream", which comes from face-to-face or online discussions between instructor and students, student monitoring software, and Web forums. This will contribute to the reformulation of teaching hypotheses, modification of software, etc. at the syllabus design level.

From the ongoing discussions, some implications for teaching translation in the future are conceivable:

- Since teaching translation into a second language is strongly related to acquisition in the second language, there should be some kind of channel for departments belonging to these two camps to readily communicate and collaborate with each other, including the sharing of resources and instructors, and the joint adventure of courses and degrees offered.
- Since information technology will play a key role in future translation profession, departments of translation should work closely with departments of computer science, establishing software development or evaluation teams or committees in order to develop and maintain software for training and adjust teaching contents and procedures.
- Computer offers an ideal environment for synthesising translation into L2 and L2

learning, especially when the use of corpora is involved. Software companies which produce MT or machine-aided translation systems like translation memories, should consider integrating reference corpora and tools into their systems, to allow trainee translators or translators on the job to learn to approximate idiomatic L2.

- The computational study of translation learners' corpora, which reveals learners' weakness in idiomatic aspect of the L2, should be implicational to researches in machine translation, since not much of this has been mentioned so far in this field. In turn, future research on collocations and conventionalised forms with computational means should produce useful tools and resources for the teaching of such.

Implications for research in second language learning can also be derived from discussions in this thesis, for example:

- Idiomatic usages, including conventionalised forms and collocations, should play a much heavier role than they do now in syllabus designing, judging from learners' inadequacies in these aspects. Grammar and vocabulary have had their share of attention, but no comprehensive listing or systematic categorising has yet been done for collocations, for example.
- Adopting some translation exercises as described in this thesis may have the benefit of forcing L2 learners to focus attention on the idiomatic aspects of the target language, and learn to discover good usages from the exploration of corpora.

Further research may include work in the following directions:

- A through study of idiomaticity from linguistic and psycholinguistic perspectives: what constitutes idiomaticity and how this can be translated across languages.
- Automatic comparison between learner corpora and native speaker corpora in respect of idiomaticity.
- How to model learners' competence in idiomaticity: how to accurately evaluate it and how to progress it systematically towards native speaker standards.
- How to design more integrated courses of translation where all modules of translation and language are interconnected in some way in terms of graded elements in both areas, offering a forest view of the distribution and progression of related modules in the entire curriculum.
- In what format and to what extent can translation be used to facilitate second

language learning.

In summary, this thesis advocates the view that to learn to translate into a second language is to a large part learning the second language itself, which, for a translator, is to a certain degree learning the idiomatic aspects of the language. Also, since learning to translate and learning a second language can conveniently converge in a modern computerised environment for instruction, it will benefit educators and trainers in the area of translation into a second language infinitely to consider designing their courses from an integrated, computational perspective.

References

Allen, J.P.B. 1975. "Some basic concepts in linguistics" In Allen, J.P.B. & Corder, S. Pit. (eds). *The Edinburgh Course in Applied Linguistics: Volume 2*: pp. 16-44. Oxford: Oxford University Press.

Altenberg, B. 1991. "Amplifier collocations in spoken English". In Johansson, S. & Stenstr"m, A. (eds). *English Computer Corpora: Selected Papers and Research Guide.* Mouton de Gruyter: Berlin.

Anderson, D. D. 1995. "Machine translation as a tool in second language learning". *CALICO*. 13(1): 68-97.

Anderson, J. 1993. "Is a communicative approach practical for teaching English in China? pros and cons". *System*. 21(4): 471-480.

Arnaudet, M. L. & Barrett, M. E. 1981. *Paragraph Development: A Guide for Students of English as a Second Language*. Englewood Cliffs, New Jersey: Prentice-Hall.

Arnold, D.J., Balkan L., Meijer S., Humphreys R.L and Sadler L . 1994. *Machine Translation: an Introductory Guide*. London: Blackwells-NCC

Auerbach, E. R . 1993. "Reexamining English only in the ESL classroom". *TESOL Quarterly*. 27(1): 9-32.

Austen, J. 1813. *Pride and Prejudice*.

Bachman, L. F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Baker, M. 1998. "Translation studies". In Baker, Mona. (ed). *Routledge Encyclopaedia of Translation Studies*. pp. 277-280. London: Routledge.

Bank of English (n.d.). Retrieved February 16, 2002, from http://www.cobuild.collins.co.uk/boe_info.html

Barnbrook, G.. 1996. "Language and Computers: A Practical Introduction to the Computer Analysis of Language". Edinburgh: Edinburgh University Press.

Bassnett, S. 1991. *Translation Studies*. London: Routledge.

Becker, J. 1975. "The phrasal lexicon". In Nash-Wabber, B. & Schank, R. (eds). *Theoretical Issues in Natural Language Processing 1*. Cambridge, Mass.: Bolt, Beranek, and Newman.

Benson, M. Benson, E. & Ilson, R. 1997. *The BBI Dictionary of English Word Combinations*. Amsterdam: John Benjamins.

Boettcher, J. 2001. "Transforming learning – Reflections on the PITAC report". *Highlights from Syllabus Magazine*. http://www.syllabus.com/syllabusmagazine/article.asp?id=5664

Bolt, P. 1993. "Grammar checking programs for learners of English as a foreign language". In Yazdani, Masoud (ed). *Multilingual Multi Media: Bridging the Language Barrier with Intelligent Systems*: 140-197. Oxford: Intellect Books.

Brooks, N. 1964. *Language and Language Learning: Theory and Practice*. 2nd ed. NY: Harcourt Brace.

Brown, G.. 1996. "Language learning, competence and performance". In Brown, Gillian. Malmkjær, Kirsten and Williams, John. (eds). *Performance & Competence in Second Language Acquisition*: pp. 187-203. Cambridge: Cambridge University Press.

Bull, S. Pain, H and Brna, P. 1993. "Student modelling in an intelligent computer assisted language learning system: the issues of language transfer and learning strategies". *International Conference on Computers in Education (Taiwan, December 1993)*.

Burns, H. L. & Capps, C. G.. 1988. "Foundations of intelligent tutoring systems: an introduction". In Polson, Martha C. & Richardson, J. Jeffrey. (eds). *Foundations of Intelligent Tutoring Systems*: 1-19. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Bussmann, H. 1996. *Routledge Dictionary of Language and Linguistics*. London: Routledge.

Caminade, M. & Pym, A. 1998. "Translator-training institutions". In Baker, Mona. (ed). *Routledge Encyclopaedia of Translation Studies*. pp. 280-285. London: Routledge.

Campbell, S. 1998. *Translation into the Second Language*. Harlow, Essex: Addison Wesley Longman.

Canale M. 1983. "From communicative competence to communicative language pedagogy". In Richards, J.C. & Schmidt, R.W. (eds). *Language and communication.* pp. 2-27. London & New York: Longman.

Canale, M. & Swain, M. 1980. "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing". *Applied Linguistics*. 1(1): 1-47.

Catford, J. C. 1965. *A Linguistic Theory of Translation*. London: Oxford University Press.

Celce-Murcia, M. 1991. "Language teaching approaches: an overview". In *Teaching English as a Second or Foreign Language*. (2nd ed.). pp. 3-11. Boston: Heinle & Heinle.

Chapelle, C. A. 1998. "Multimedia CALL: lessons to be learned from research on instructed SLA". *Language Learning & Technology*. 2(1): 22-34.

Chen, H-J. H. 1999. "Creating a virtual language lab: an EFL experience at National Taiwan Ocean University". *ReCALL*. 15(2): 20-30.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass: MIT Press

Cobb, T. 1999. "Breadth and depth of lexical acquisition with hands-on concordancing". *CALL*. 12(4): 345-360.

Connell, T. 1999. "Web support for distance learning in the field of translation". *ReCALL*. 11(2): 31-37.

Cook, G. 1998. "Use of translation in language teaching". In Baker, Mona. (ed). *Routledge Encyclopaedia of Translation Studies*. pp. 277-280. London: Routledge.

Cook, V. J. 1996. "Competence and multi-competence". In Brown, Gillian, Malmkjær, Kirsten and Williams, John. (eds). *Performance & Competence in Second Language Acquisition*: pp. 57-69. Cambridge: Cambridge University Press.

Crawford-Lange, L. M. 1987. "Curricular alternatives for second-language learning". In Long, M.H. and Richards, J.C. (eds). *Methodology in TESOL: A Book of Readings.* pp. 120-144. Singapore: Harper & Row Publishers, Asia, Pte., Ltd.

Croft, K. 1972. "Trends and practices". In Croft, Kenneth. (ed). *Readings on English as a Second Language*. pp. 1-12. Cambridge, Mass: Winthrop Publishers.

Cruse D.A. 1986. *Lexical Semantics*. Cambridge University Press: Cambridge.

Curtis, S. A, Duchastel, Joanne & Radic, Nebojsa. 1999. "Proposal for an online language course". *ReCALL*. 11(2): 38-45.

DeCesaris, J. N. 1996. "Computerized translation managers as teaching aids". In Dollerup, Cay & Appel, Vibeke. (eds). *Teaching Translation and Interpreting 3: New Horizons*: Papers from the Third Language International Conference, Elsinore, Denmark, 9-11 June 1995. Amsterdam: John Benjamins.

Dirksen, C. 1990. "Learning styles of Mainland Chinese students of English". *IDEAL*. 5: 29-38.

Dodds, J. M. 1999. "Friends, false friends and foes or back to basics in L1 to L2". In Anderman, Gunilla & Rogers, Margaret (eds). *Word, Text, Translation: Liber Amicorum for Peter Newmark*: 56-65. Clevedon: Multilingual Matters.

Douglas, S. A. 1995. "LingWorlds: an intelligent object-oriented environment for second language tutoring". In Holland, V.M., Kaplan, J.D. & Sams, M.R. (eds). *Intelligent Language Tutors: Theory Shaping Technology*: 201-220. Mahwah, NJ: Lawrence Erlbaum Associates.

Duff, A. *Translation*. 1989. Oxford: Oxford University Press.

Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Fawcett, P. 1998. "Linguistic approaches". In Baker, Mona. (ed). *Routledge Encyclopaedia of Translation Studies*. 120-125. London: Routledge.

Finocchiaro, M. & Brumfit, C. 1983. *The Functional-Notional Approach: From Theory to Practice*. NY: Oxford University Press.

Fleta, B. M., Sabater, C.P., Salom, L.G., Guillot, C.P., Monreal, C.S. & Turney, E. 1999. "Evaluating multimedia programs for language learning: a case study". *ReCALL*. 11(3): 50-57.

Frederiksen, C. H.; Donin, J. & Decary, M. "A discourse processing approach to computer-assisted language learning". In Holland, V.M., Kaplan, J.D. & Sams, M.R. (eds). *Intelligent Language Tutors: Theory Shaping Technology*: 99-120. Mahwah, NJ: Lawrence Erlbaum Associates.

Freigang, K-H. 1998. "Machine-aided translation". In Baker, Mona. (ed). *Routledge Encyclopaedia of Translation Studies*. pp. 134-136. London: Routledge.

Garside, R, Leech, G. & McEnery, T. 1997. *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London: Longman.

Gavioli, L. 1997. "Exploring tests through the concordancer: guiding the learner". In Wichmann, Anne; Fligelstone, Steven; McEnery, Tony & Knowles, Gerry. (eds). *Teaching and Language Corpora*: 83-99. London: Longman.

Goodfellow, R. & Metcalfe, P. 1997. "The challenge - back to basics or brave new world?". *ReCALL*. 9(2): 4-7.

Gitsaki, C. 1996. "The development of ESL collocational knowledge". PhD thesis. Brisbane, Australia: Center for Language Teaching and Research, The University of Queensland.

Granger, S. 1996. "Learner English around the world". In Greenbaum, Sidney. (ed). *Comparing English Worldwide: The International Corpus of English*. pp. 13-24. Oxford: Clarendon Press.

Graves, K. 2000. *Designing Language Courses: A Guide for Teachers*. Boston: Heinle & Heinle.

Gross, A. 1992. "Limitations of computers as translation tools". In Newton, John. (ed). *Computers in Translation: A Practical Appraisal*: pp. 96-130. London: Routledge.

Harben, P. 1999. "An exercise in applying pedagogical principles to multimedia CALL materials design". *ReCALL*. 11(3): 25-33.

Harley, T. A. 1995. *The Psychology of Language: From Data to Theory*. Hove, East Sussex: Erlbaum.

Harris, T. & Rowe, A. 1997. *Exploring English 5*. NY: Addison Wesley Longman.

Hatim, Basil & Mason, I. 1990. *Discourse and the Translator*. N.Y.: Addison Wesley Longman Inc.

Higgins, J. 1995. *Computers and English Language Learning*. Oxford: Intellect Ltd.

Hoey, M. & Houghton, D. 1998. "Contrastive analysis and translation". In Baker, Mona. (ed). *Routledge Encyclopaedia of Translation Studies*. London: Routledge

Howarth, P. 1998. "The phraseology of learners' academic writing". In Cowie, A.P. (ed). *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.

Howatt, A.P.R. 1984. *A History of English Language Teaching*. Oxford: Oxford University Press.

Hubbard, P. 1992. "A methodological framework for CALL courseware development". In Pennington, Martha C. and Stevens, Vance. (eds). *Computers in Applied Linguistics: An International Perspective*: 39-65. Clevedon: Multilingual Matters.

Hubbard, P. L. 1996. "Elements of CALL methodology: development, evaluation, and implementation". In Pennington, Martha C. (ed). *The Power of CALL*: 15-32. Houston: Athelstan Publications.

Hung, E. 1996. "Translation curricula development in Chinese communities". In Dollerup, Cay & Appel, Vibeke (eds). *Teaching Translation and Interpreting 3: New*

Horizons: Papers from the Third International Conference, Elsinore, Denmark, 9-11 June 1995. pp. 31-44. Amsterdam: John Benjamins.

Hutchins, J. 1995. "Reflections on the history and present state of machine translation". In: *MT Summit V proceedings,* Luxembourg, July 10-13, 1995: pp. 89-965.

Hutchins, J. 1997. "Translation technology and the translator". In Greensmith, Catherine & Vandamme, Marilyn (eds). *Proceedings of the Eleventh Conferences of the Institute of Translation and Interpreting*, 8-10 May 1997, London: pp. 113-120.

Hutchins, J. 1998. "Twenty years of translating and the computer". *Proceedings from the Aslib conference held on 12 & 13 November 1998*, London.

Hutchins, J. 1999. "Retrospect and prospect in computer-based translation". In *Machine Translation Summit VII*, 13th-17th September 1999: pp.30-34.

Johns, T. 1997. "Contexts: the Background, Development and Trialling of a Concordance-based CALL Program". In Wichmann, Anne; Fligelstone, Steven; McEnery, Tony & Knowles, Gerry. (eds). *Teaching and Language Corpora*: 100-115. London: Longman.

Jung, E. H. (Sarah). 2000. "Focus on Grammar: Intermediate Level". *CALICO Software Review.* Available from: http://astro.temple.edu/~jburston/CALICO/review/focusgram00.htm.

Kenning, M.M. 1990. *Computers and Language Learning: Current Theory and Practice*. Chichester, West Sussex: Ellis Horwood.

Kenny, D. 1998. "Corpora in translation studies". In Baker, Mona. (ed). *Routledge Encyclopaedia of Translation Studies*. 50-53. London: Routledge

Krouglov, A. 1996. "Social and cultural differences". In Dollerup, Cay & Appel, Vibeke (eds). *Teaching Translation and Interpreting 3: New Horizons: Papers from the Third International Conference, Elsinore, Denmark, 9-11 June 1995*. pp. 81-87. Amsterdam: John Benjamins.

Kumaravadivelu, B. 1994. "The postmethod condition: (e)merging strategies for second/foreign language teaching". *TESOL Quarterly*. 28(1): 27-48.

Lai, C. 1994. "Communication failure in the language classroom: an exploration of causes". *RELC Journal*. 25(1): 99-129.

Levelt, W.J.M. 1989. *Speaking: From Intention to Articulation*. Cambridge, Mass.: The MIT Press:

Lian, A. 1992. "Intelligence in computer-aided language learning". In Pennington, Martha C. & Stevens, Vance. (eds). *Computers in Applied Linguistics: An International Perspective*: 66-154. Clevedon, Avon: Multilingual Matters.

Lindquist, H. 1999. "Electronic corpora as tools for translation". In Anderman, Gunilla & Rogers, Margaret (eds). *Word, Text, Translation: Liber Amicorum for Peter Newmark*: pp.179-189. Clevedon: Multilingual Matters

Liu, M. 1997. *A Workbook for English-Chinese Translation* (in Chinese). Taipei: Bookman.

Long, M.H. & Crookes, G.. 1992. "Three approaches to task-based syllabus design". *TESOL Quarterly*. 26(1): 27-50.

Lyons, J. 1996. "On competence and performance and related notions". In Brown, Gillian. Malmkjær, Kirsten and Williams, John. (eds). *Performance & Competence in Second Language Acquisition*: pp. 11-32. Cambridge: Cambridge University Press.

Manning, C. D. & Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, Mass: The MIT Press.

Markee, N. 1997. *Managing Curricular Innovation*. Cambridge: Cambridge University Press.

McCormack, C. & Jones, D. 1998. *Building a Web-Based Education System*. Boston: John Wiley & Sons.

McDonough, J. & Shaw, C. 1993. *Materials and Methods in ELT*. Oxford: Blackwell Publishers.

McEnery, T. & Wilson, A. 1997. "Teaching and language corpora (TALC)". *ReCALL*.

9(1): 5-14.

Melby, A. 1992. "The translator workstation". In Newton, John. (ed). *Computers in Translation: A Practical Appraisal*. pp. 147-165. London: Routledge.

Melby, A. K. 1994. "The translator workstation". In Hammond, Deanna L. (ed). *Professional Issues for Translators and Interpreters: American Translators Association Scholarly Monograph Series, Volume VII.* pp. 127-147. Amsterdam: John Benjamins.

Meskill, C. 1996. "Computers, creativity and communicative competence: an association machine". *CALL*. 9(2-3):115-123.

Mills, H. 1982. *Connecting and Combining*. Glenview, Illinois: Scott, Foresman and Company

Murison-Bowie, S. 1996. "Linguistic corpora and language teaching". *Annual Review of Applied Linguistics* 16: 182-199.

Murray, J. H. 1995. "Lessons learned from the Athena language learning project: using natural language processing, graphics, speech processing, and interactive video for communication-based language learning". In Holland, V.M., Kaplan, J.D. & Sams, M.R. (eds). *Intelligent Language Tutors: Theory Shaping Technology*: 243-256. Mahwah, NJ: Lawrence Erlbaum Associates.

Nadell, J., McMeniman, L. & Langan, J. 1997. *The Macmillan Writer: Rhetoric and Reader* (3rd ed). Boston: Allyn and Bacon.

Nattinger, J. R. & DeCarrico, J. S. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Neubert, A. & Shreve, G. M. 1992. *Translation as Text*. Kent, Ohio: The Kent State University Press.

Newmark, P. 1991. *About Translation*. Clevedon: Multilingual Matters.

Newmark, P. 1988. *A Textbook of Translation*. Hemel Hempstead, Hertfordshire: Prentice Hall.

Nida, E. A. 1964. *Towards a science of translating*. Leiden: E. J. Brill.

Nida, E. A. 1997. "Translation in the information age". In Labrum, Marian B. (ed). *The Changing Scene in World Languages: Issues and Challenges: American Translators Association Scholarly Monograph Series, Volume IX*. pp. 9-17.

Nunan, D. 1989. *Designing Tasks for the Communicative Classroom*. Cambridge: Cambridge University Press.

Nunan, D. 1999. "Yes, But Is It English?". Originally in *TESOL Matters*, December/January 1999. Retrieved 14 February, 2002, from: http://www.tesolgreece.com/yesbut.html.

Nunan, D. 2001. "Aspects of task-based syllabus design". *Karen's Linguistic Issues*. December 2001. Retrieved 14 February, 2002, from: http://www3.bc.sympatico.ca/linguisticsissues/syllabusdesign.html.

Oakes, M. P. 1998. *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Ocrat Chinese Pages (n.d.). Retrieved 14 March, 2002, from: http://www.ocrat.com/ocrat/biling/pridprej/

O'Hagan, M. 1996. *The Coming Industry of Teletranslation*. Clevedon: Multilingual Matters.

Pain, H. 2002. *Personal Correspondence*.

Pawley, A. & Syder, E. 1983. "Two puzzles for linguistic theory: nativelike selection and nativelike fluency". In J.C. Richards and R. Schmidt (eds): *Language and Communication*. London: Longman.

Pennington, M. C. 1996. "The power of the computer in language education". In Pennington, Martha C. (ed). *The Power of CALL*: 1-14. Houston: Athelstan Publications.

Pica, T. 1994. "Questions from the language classroom: research perspectives". *TESOL Quarterly*. 28(1): 49-77.

Prabhu, N.S. 1987. *Second Language Pedagogy.* Oxford: Oxford University Press.

Purgason, K. B. 1991. "Planning lessons and units". In Celce-Murcia (ed). *Teaching English as a Second or Foreign Language* (Second Edition). pp.419-431. Boston: Heinle & Heinle.

Raimes, A. 1983. *Techniques in Teaching Writing.* Oxford: Oxford University Press

Reynolds, R. 2001. "Translation instruction on-line". Paper presented at the Fifth Annual Conference on Translation Instruction, National Taiwan Normal University, January 6, 2001.

Richards, J. C. & Lockhart, C. 1994. *Reflective Teaching in Second Language Classrooms.* Cambridge: Cambridge University Press.

Richards, J. C. & Rodgers, T. S. 1986. *Approaches and Methods in Language Teaching: A Description and Analysis.* Cambridge: Cambridge University Press.

Rodgers, T. S. 1989. "Syllabus design, curriculum development and polity determination". In Johnson, R.K. (ed). *The Second Language Curriculum.* pp. 24-34. Cambridge: Cambridge University Press.

Ruetten, M. K. 1997. *Developing Composition Skills: Rhetoric and Grammar.* Boston: Heinle & Heinle.

Schmidt, R. W. 1990. "The role of consciousness in second language learning". *Applied Linguistics.* 11(2): 129-158.

Schuster, E. 1996. "The role of native grammars in correcting errors in second language learning". *Computational Intelligence.* 2: 93-98.

Selinker, L. 1969. "Language Transfer". *General Linguistics.* 9(2): 67-92.

Sentence, S. 1997. "A Rule Network For English Article usage within an Intelligent Language Tutoring System". *CALL.* 10(2): 173-200.

Sheely, S., Veness, D. & Rankine, L. 2001. "Building the Web Interactive Study Environment: Mainstreaming online teaching and learning at the University of Western Sydney". *Australian Journal of Educational Technology.* 17(1): 80-95.

Shei, C-C. 2001. "FollowYou!: An Automatic Language Lesson Generation System". *Computer Assisted Language Learning* (CALL). 14(2): 129-144.

Shei, C-C. & Pain, H. 2000. "An ESL Writer's Collocational Aid". *Computer Assisted Language Learning* (CALL). 13(2): 167-182.

Shei, C-C. & Pain, H. 2001. "Learning a Foreign Language Through Machine Translation: Focusing on Sentence Stems and Collocations". *AI-ED Workshop on CALL: Implementing Intelligent Language Tutoring Systems.* May 19-23, 2001. San Antonio, US.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. & Renouf, A. 1988. "A lexical syllabus for language learning". In Carter, R. & McCarthy, M. (eds). *Vocabulary and Language Teaching*. pp. 141-160. Harlow, Essex: Longman.

Skehan, P. 1998. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Smadja, F. 1993. "Retrieving collocations from text: Xtract". In Susan Armstrong. (ed). *Using Large Corpora.* Cambridge, Mass: MIT Press.

Sofer, M. 1998. *The Translator's Handbook* (2nd ed). Rockville, Maryland: Schreiber Publishing, Inc.

Somers, H. L. 1998a. "Machine translation: applications". In Baker, Mona. (ed). *Routledge Encyclopaedia of Translation Studies*. pp. 136-139. London: Routledge.

Somers, H. L. 1998b. "Machine translation: history". In Baker, Mona. (ed). *Routledge Encyclopaedia of Translation Studies*. pp. 140-143. London: Routledge.

Somers, H. L. 1998c. "Machine translation: methodology". In Baker, Mona. (ed). *Routledge Encyclopaedia of Translation Studies*. pp. 143-149. London: Routledge.

Spolsky, B. 1972. "What does it mean to know a language?" In Croft, Kenneth. (ed).

Readings on English as a Second Language. pp. 26-42. Cambridge, Mass: Winthrop Publishers.

Stern, H.H. 1992. *Issues and Options in Language Teaching*. Edited by Patrick Allen and Birgit Harley. Oxford: Oxford University Press.

Stevens, V. 1995. "Concordancing with Language Learners: Why? When? What?". *CAELL Journal* 6(2): 2-10.

Strevens, P. D. 1987. "The Nature of Language Teaching". In Long, M.H. and Richards, J.C. (eds). *Methodology in TESOL: A Book of Readings*. pp. 10-25. Singapore: Harper & Row Publishers, Asia, Pte., Ltd.

Stubbs, M. 1995. "Corpus evidence for norms of lexical collocation". In Cook, G. & Seidlhofer, B. (eds). *Principle and Practice in Applied Linguistics: Studies in honour of H.G.Widdowson.* Oxford: Oxford University Press.

Swain, M. 1985. "Communicative competence: Some roles of comprehensible output in its development". In Gass., S.M. & Madden, C.G. (eds). *Input in Second Language Acquisition*. pp. 235-253. Boston: Heinle & Heinle.

Talbot, G. 1996. "Looking up in anger: translation practice in the CALL lab". *ReCALL*. 8(1): 20-23.

Tarvin, W. L. & Al-Arishi, A. Y. 1991. "Rethinking communicative language teaching: reflection and the EFL classroom". *TESOL Quarterly*. 25(1): 9-27.

TellMeMore. (n.d). Retrieved February 2, 2002, from http://www.owncampus.com/tellmemore

Tomlin, R. S. 1995. "Modeling individual tutorial interactions: theoretical and empirical bases of ICALL". In Holland, V.M., Kaplan, J.D. & Sams, M.R. (eds). *Intelligent Language Tutors: Theory Shaping Technology*: 221-242. Mahwah, NJ: Lawrence Erlbaum Associates.

Triola, M. F. 1998. *Elementary Statistics* (7th ed). Reading, Massachusetts: Addison-Wesley.

Trujillo, A. 1999. *Translation Engines: Techniques for Machine Translation*. London:

Springer-Verlag.

Ulrych, M. 1996. "Real-world criteria in translation pedagogy". In Dollerup, Cay & Appel, Vibeke (eds). *Teaching Translation and Interpreting 3: New Horizons: Papers from the Third International Conference, Elsinore, Denmark, 9-11 June 1995*. pp. 31-44. Amsterdam: John Benjamins.

van Lier, L. 1988. *The classroom and the Language Learner. : Ethnography and Second-language Classroom Research*. London: Longman.

Vermeer, H. J. 1998. "Didactics of translation". In Baker, Mona. (ed). *Routledge Encyclopaedia of Translation Studies*. pp. 60-63. London: Routledge.

Wältermann, D. 1994. "Machine translation systems in a translation curriculum". In Dollerup, Cay & Lindegaard, Annette (eds). *Teaching Translation and Interpreting 2: Insights, Aims, Visions*: 309-317. Amsterdam: John Benjamins.

Wang, L. 2001. "Exploring parallel concordancing in English and Chinese". *Language Learning & Technology*. 5(3): 174-184. Available from: http://llt.msu.edu/vol5num3/wang/ .

Wang, Yang & Garigliano, R. 1992. "An intelligent language tutoring system for handling errors caused by transfer". *Intelligent Tutoring Systems: Second International Conference,* ITS '92, Montreal, Canada, Jun3 10-12, 1992.

Whistle, J. 1999. "Concordancing with students using an 'off-the-Web' corpus". *ReCALL*. 11(2): 74-80.

White, R. & Arndt, V. 1991. *Process Writing*. London: Longman.

Widdowson, H.G. 1989. "Knowledge of language and ability for use". *Applied Linguistics*. 10(2): 128-137.

Wilkins, D.A. 1976. *Notional Syllabuses*. Oxford: Oxford University Press.

Wolfson, N. 1989. *Perspectives: Sociolinguistics and TESOL*. Boston: Heinle & Heinle.

WordNet, (n.d). Retrieved 14 March, 2002, from: http://www.cogsci.princeton.edu/~wn/

Wu, Q-C. 1982. *Chinese-English Translation Through Contrastive Analysis* (in Chinese). Taipei: Crane

Yalden, J. 1987. *The Communicative Syllabus: Evolution, Design and Implementation.* Englewood Cliffs, New Jersey: Prentice-Hall.

Yang, J. & Lange, E. D. 1998. "SYSTRAN on AltaVista: a user study on real-time machine translation on the Internet". Retrieved 7 May, 2001, from: http://www.systransoft.com/Papers/ppr_alta.htm .

Zanettin, F. 1998. "Bilingual comparable corpora and the training of translators". *Meta*. XLIII(4): 1-14.

Zhang, P., Yu, Y., Li, Z. & Peng, M. 1993. *A Course in Translation* (in Chinese). Taipei: Bookman.

Zhou, Z. 1996. *First Steps in Translation* (in Chinese). Hong Kong: Bookman Books.

Appendix A

Detailed marking of Idiomatic scores

1. *designed specifically*

Students' variant	Number of subjects	Score
designed especially	3	8
designed specially	2	8
especially designed	4	7
specially designed	9	7
particularly designed	2	6
only designed	1	5
designed	19	3
especially set	1	2
produce	1	1
Total	42	

2. *preschool children*

Students' variant	Number of subjects	Score
preschool children	21	10
pre-school-aged children	1	8
preschoolers	4	7
preschool babies	1	6
children of preschool age	1	5
children before school age	6	4
children under the school age	1	3
children who are before school age	2	2
before school age children	1	1
infant before school age	1	1
infant children before school age	1	1
those babies before going to school	1	1
kids whose ages under 7	1	1

Total	42	

3. *look like*

Students' variant	**Number of subjects**	**Score**
look like	27	10
look very like	2	9
seem like	2	8
are like	6	8
is like	1	7
are very similar to	1	5
are look like	1	4
are like to	1	3
(not translated)	1	0
Total	42	

4. *communicate with*

Students' variant	**Number of subjects**	**Score**
communicate with	14	10
communication between	1	8
connect with	2	7
establish contact with	1	6
make contact with	1	6
connect to	1	5
get in touch with	1	4
get into contact with	1	3
transmit information with	1	3
contact	2	3
is connection with	1	2
is link with	1	2
is connection between	2	2
is conjunction between	1	1
is connection	1	1
connect	1	1
link	1	1
contact with[76]	5	1
is via between[77]	1	0
(not translated)	3	0

[76] *Contact* incorrectly used as an intransitive verb.
[77] *Via* incorrectly used as a noun.

Total	42

5. *the rest of the world*

Students' variant	**Number of subjects**	**Score**
the outside world	12	8
the external world	4	8
the world outside	1	7
the world	2	6
the outside	11	5
other worlds	1	4
other people	1	3
others	4	2
outsiders	1	2
outer world	1	1
the outside people	1	1
the public	1	1
outer rooms	1	1
(not translated)	1	0
Total	42	

6. *thanks to*

Students' variant	**Number of subjects**	**Score**
by way of	2	8
by means of	1	8
through	25	6
therefore	1	5
by	9	4
also	1	2
over	1	1
getting through	1	1
by through	1	1
Total	42	

7. *interact with*

Students' variant	Number of subjects	Score
interact with	15	10
communicate with	4	8
connect with	1	7
play with	5	6
have mutual activities with	2	5
connect mutually with	1	4
act with	1	3
have action together with	1	2
go along with	1	1
be in touch with	1	1
exchange with	1	1
interact on	1	1
interaction with[78]	1	1
talk to each other	1	1
produce mutuality	1	1
share	1	1
(not translated)	4	0
Total	42	

8. *participate in*

Students' variant	Number of subjects	Score
participate in	19	10
take part in	2	9
join in	2	8
join[79]	14	3
play ... with	2	2
attend	1	1
participate	2	1
Total	42	

[78] *Interaction* used as a verb.

[79] As the objects here are 'activities' rather than people or organisation, *join in* is acceptable but not *join*.

9. *no matter whether*

Students' variant	**Number of subjects**	**Score**
no matter what	2	8
no matter when	1	8
no matter while	1	8
no matter how	1	8
no matter	13	6
whether … or	8	6
either … or	1	5
regardless of	5	4
not only … but also	2	3
whatever	3	2
whenever	1	1
(not translated)	4	0
Total	42	

10. *bear remarkable resemblance*

Students' variant	**Number of subjects**	**Score**
resemble	2	8
look like	1	7
be like	10	6
be similar to	24	5
be the same as	2	4
be the same with	1	1
be similar with	1	1
match together	1	1
Total	42	

Appendix B

A Questionnaire About English Collocation

==========

I am a:

- ( )native speaker of English
- ( )non-native speaker of English -- optional: My native language is:( )

==========

Instruction

Please choose the words you would accept in each sentence and rank them, focusing on the compatibility between the verb-noun pair. For example:

- I can not _____ any responsibility for what happens.
  (2)accept (3)assume ( )get (1)take

Here "take" is considered best to go with "responsibility", "accept" second best, "assume" the third, and "get" unacceptable. Please remember, (1) is for the best, and (2) the second best etc., not the other way round. You can rank all four, or leave blank the ones you find unacceptable.

==========

1. He said he had already _____ a complaint.
   ( )lodged ( )made ( )reported ( )submitted
2. We usually _____ an activity with a goal in mind.
   ( )do ( )execute ( )perform ( )practise
3. This paper will _____ the issue in the context of community work.
   ( )address ( )cover ( )examine ( )treat
4. In this way students _____ confidence in using the library and have the opportunity to meet friends.
   ( )acquire ( )gain ( )get ( )obtain
5. The primary intention of the speaker is to _____ information.
   ( )convey ( )deliver ( )transfer ( )transmit
6. The clinical teacher must _____ a good relationship with the learner.
   ( )build ( )construct ( )establish ( )make
7. You're obviously an experienced speaker; you know how to _____ a response.
   ( )arouse ( )elicit ( )extract ( )stimulate
8. There are people out there who will use any excuse to _____ trouble.
   ( )cause ( )generate ( )make ( )stir up

9. It is astounding that people managed to _____ agreement so quickly.
   ( )achieve ( )get ( )make ( )reach
10. They agreed to _____ research on land-based disposal.
   ( )do ( )conduct ( )perform ( )undertake
11. She would treat this as an unexpected opportunity to _____ experience in mass-production fashion.
   ( )acquire ( )get ( )gain ( )obtain
12. I will _____ an example where all the methods can be used.
   ( )cite ( )make ( )quote ( )take
13. The role of the Architect is to _____ such a service.
   ( )conduct ( )execute ( )perform ( )provide
14. They practise dissection in order to _____ knowledge of the human body.
   ( )absorb ( )acquire ( )learn ( )obtain
15. A sunny position and well-drained soil _____ success.
   ( )assure ( )ensure ( )guarantee ( )warrant

==========

Appendix C

Scores for the Questionnaire

Chinese group	Native speaker group	European group
129	169	53
96	170	156
140	171	165
139	194	111
145	174	161
120	176	128
110	148	145
139	169	156
159	154	162
136	162	147
145	173	137
113	173	141
128	162	162
151	170	154
108	111	149
177	183	153
139	142	97
125	163	174
163	135	147
122	146	132
149	153	143
110	186	137
92	186	154
108	166	159
91	138	160
76	159	141
112	156	126
99	174	110
136	182	131
60	164	175
131	182	168
123	131	170
102	156	

113	143	
100	158	
123	162	
103	144	
134	179	
111	141	
133	165	
125	114	
119	126	
	135	
	181	
	166	
	146	
	139	
	141	
	178	
	145	
	162	
	159	
	174	
	176	
	150	
	152	
	135	
	167	
	136	
	169	
	177	
	164	
	194	
	164	
	179	
	157	
	176	
	158	
	162	
	173	
	182	

	164 150 178 184 162 167 155 163 180 165 169 180 165 166 176 207 192 193 149 154 163 172 177 159 140 178 192 187 167 191 172 194 198 102 125 171 167 165	

	175 186 208 132 183 145 173 184 168 125	
42 subjects	119 subjects	32 subjects

Appendix D

Full Lists of Courses Offered in the Department of Translation, CJU

Semester One

First year		
Module	Credit	Weekly hour
English conversation and viewing/listening (I)	2	2
English grammar and writing (I)	2	2
General Translation: L2-L1 (I)	2	2
Mandarin (I)	2	2
Philosophy of life	3	3
Introduction to the computer (I)	3	3
Modern foreign language: Japanese (I)	2	2
Modern foreign language: German (I)	2	2
Modern foreign language: French (I)	2	2
Modern foreign language conversation, elementary: Japanese (I)	2	2
Modern foreign language conversation, elementary: German (I)	2	2
Modern foreign language conversation, elementary: French (I)	2	2
Anthology of contemporary Western literature	2	2
Learning from service (I)	1	1
Military training (I)	0	2
Physical education (I)	0	2
English (I)	2	2
English listening comprehension (I)	1	1

Second year		
Module	Credit	Weekly hour
English conversation and viewing/listening (III)	2	2
English grammar and writing (III)	2	2
General Translation (III)	2	2
English for economics and trading	2	2
Sight translation: L2-L1 (I)	2	2
Introduction to the computer (I)	2	2
Modern foreign language: Japanese (I)	2	2
Modern foreign language: German (I)	2	2
Modern foreign language: French (I)	2	2
Modern foreign language conversation, lower intermediate: Japanese (I)	2	2
Modern foreign language conversation, lower intermediate: German (I)	2	2
Modern foreign language conversation, lower intermediate: French (I)	2	2
Introduction to broadcasting	2	2
Politics	2	2
News English	2	2
Military training (III)	1	2
Physical education (III)	0	2
English (III)	2	2
Common knowledge courses	2	2

Third year		
Module	Credit	Weekly hour
English conversation (I)	2	2
English grammar and writing (V)	2	2
Advanced Translation: L1-L2	2	2
Advanced Translation: L2-L1	2	2
Sight translation: L1-L2	2	2
General consecutive interpreting: L1-L2 (I)	2	2
Advanced consecutive interpreting: L1-L2 (I)	2	2
Advanced consecutive interpreting: L2-L1 (I)	2	2
English listening and speaking practical (I)	2	2
Modern foreign language reading and writing,: Japanese (I)	2	2
Modern foreign language reading and writing,: German (I)	2	2
Modern foreign language reading and writing,: French (I)	2	2
Modern foreign language, intermediate: Japanese (I)	2	2
Modern foreign language, intermediate: German (I)	2	2
Modern foreign language, intermediate: French (I)	2	2
Modern foreign language conversation, intermediate: Japanese (I)	2	2
Modern foreign language conversation, intermediate: German (I)	2	2
Modern foreign language conversation, intermediate: French (I)	2	2
Mandarin syntax	2	2
Introduction to English and American literature	2	2
Elementary medical science	2	2
Computer and translation	2	2
Physical education (V)	0	2
Common knowledge courses	2	2

Fourth year		
Module	Credit	Weekly hour
Academic paper writing	2	2
Professional translation: L1-L2 (I)	2	2
Professional translation: L2-L1 (I)	2	2
Advanced English writing (III)	2	2
English listening and speaking practical (III)	2	2
Advanced consecutive interpreting: L1-L2	2	2
General simultaneous interpreting: L2-L1	2	2
Advanced consecutive interpreting: L2-L1 (I)	2	2
Modern foreign language, advanced: Japanese (I)	2	2
Modern foreign language, advanced: German (I)	2	2
Modern foreign language, advanced: French (I)	2	2
Modern foreign language conversation, advanced: Japanese (I)	2	2
Modern foreign language conversation, advanced: German (I)	2	2
Modern foreign language conversation, advanced: French (I)	2	2
General translation into second foreign language: Japanese (I)	2	2
General translation into second foreign language: German (I)	2	2
General translation into second foreign language: French (I)	2	2
Introduction to English and American literature	2	2
Physical education (VII)	0	2
Common knowledge courses	2	2

Semester Two

First year		
Module	Credit	Weekly hour
English conversation and viewing/listening (II)	2	2
English grammar and writing (II)	2	2
General Translation: L1-L2 (I)	2	2
Mandarin (II)	2	2
Philosophy of life	3	3
Introduction to the computer (II)	3	3
Modern foreign language: Japanese (II)	2	2
Modern foreign language: German (II)	2	2
Modern foreign language: French (II)	2	2
Modern foreign language conversation, elementary: Japanese (II)	2	2
Modern foreign language conversation, elementary: German (II)	2	2
Modern foreign language conversation, elementary: French (II)	2	2
Anthology of contemporary Western literature	2	2
Learning from service (II)	1	1
Military training (II)	0	2
Physical education (II)	0	2
English (II)	2	2
English listening comprehension (II)	1	1

Second year		
Module	Credit	Weekly hour
English conversation and viewing/listening (IV)	2	2
English grammar and writing (IV)	2	2
General Translation: L2-L1 (II)	2	2
English for science and technologies	2	2
Sight translation: L2-L1 (II)	2	2
General consecutive interpreting: L2-L1	2	2
Modern foreign language: Japanese (II)	2	2
Modern foreign language: German (II)	2	2
Modern foreign language: French (II)	2	2
Modern foreign language conversation, lower intermediate: Japanese (II)	2	2
Modern foreign language conversation, lower intermediate: German (II)	2	2
Modern foreign language conversation, lower intermediate: French (II)	2	2
ELT methodologies	2	2
History of Taiwanese foreign affairs	2	2
International relationship	2	2
Introduction to management	2	2
Military training (II)	1	2
Physical education (IV)	0	2
English (IV)	2	2
Common knowledge courses	2	2

Third year		
Module	Credit	Weekly hour
Advanced English writing (II)	2	2
Advanced Translation: L2-L1 (II)	2	2
Advanced Translation: L1-L2 (II)	2	2
English conversation (II)	2	2
Film and television translation: L2-L1	2	2
Sight translation: L1-L2	2	2
General consecutive interpreting: L1-L2 (I)	2	2
Advanced consecutive interpreting: L1-L2 (II)	2	2
Advanced consecutive interpreting: L2-L1 (II)	2	2
English listening and speaking practical (II)	2	2
Modern foreign language reading and writing,: Japanese (II)	2	2
Modern foreign language reading and writing,: German (II)	2	2
Modern foreign language reading and writing,: French (II)	2	2
Modern foreign language, intermediate: Japanese (II)	2	2
Modern foreign language, intermediate: German (II)	2	2
Modern foreign language, intermediate: French (II)	2	2
Modern foreign language conversation, intermediate: Japanese (II)	2	2
Modern foreign language conversation, intermediate: German (II)	2	2
Modern foreign language conversation, intermediate: French (II)	2	2
Mandarin syntax	2	2
Introduction to English and American literature	2	2
Elementary medical science	2	2
Physical education (VI)	0	2
Common knowledge courses	2	2

Fourth year		
Module	Credit	Weekly hour
Advanced science and technology translation: L1-L2	2	2
Professional translation: L2-L1 (I)	2	2
Graduation project	2	2
English listening and speaking practical (IV)	2	2
General simultaneous interpreting: L1-L2	2	2
Advanced consecutive interpreting: L2-L1 (I)	2	2
Modern foreign language, advanced: Japanese (I)	2	2
Modern foreign language, advanced: German (II)	2	2
Modern foreign language, advanced: French (II)	2	2
Modern foreign language conversation, advanced: Japanese (II)	2	2
Modern foreign language conversation, advanced: German (II)	2	2
Modern foreign language conversation, advanced: French (II)	2	2
Children's literature	2	2
Physical education (VIII)	0	2
Common knowledge courses	2	2

Appendix E

Related Publications

1. **An ESL Writer' Collocational Aid**
2. **Learning a Foreign Language Through Machine Translation: Focusing on Sentence Stems and Collocations**
3. **Computational Approach to the Teaching of Translation Methods**